

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2002, 14:01:57 ; Search time 35.72 Seconds
(without alignments)
3521.177 Million cell updates/sec

Title: US-09-468-147-91

Perfect score: 9013

Sequence: 1 PGITTAIEQAALAAANSALA.....FTETIKPVLDITNSIQRAVE 1698

Scoring table:

BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :

A_Geneseq.1101.*
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9001	99.9	1698	20	AAV31381
2	8812.5	97.8	1708	20	AAW93408
3	8753.5	97.1	1708	20	AAV31384
4	7456	82.7	1693	17	AAV31813
5	7448	82.6	1693	19	AAW80196
6	7448	82.6	1693	21	AAW824119
7	7448	82.6	1693	22	AAW62522
8	7439	82.5	1693	19	AAW81519
9	7437	82.5	1693	19	AAW76368
10	7437	82.5	1693	19	AAW71209
11	7436	82.5	1693	15	AAW51264

12	7400	82.1	1693	12	AAV14618
13	2057	22.8	431	12	AAV14615
14	2057	22.8	431	19	AAW80195
15	2057	22.8	431	19	AAW71208
16	2057	22.8	431	21	AAW24118
17	2057	22.8	431	22	AAW62521
18	705	7.8	138	20	AAV31406
19	701	7.8	138	20	AAV31405
20	504	5.6	1704	15	AAW49657
21	403	4.5	2115	19	AAW59276
22	399	4.4	1645	22	AAV31997
23	393.5	4.4	1644	22	AAW32000
24	374.5	4.2	2205	16	AAW79048
25	360	4.0	1233	19	AAW44727
26	360	4.0	1233	19	AAW41935
27	268.5	3.0	1144	22	AAW31995
28	259	2.9	1143	22	AAW31998
29	213.5	2.4	2161	20	AAW87724
30	185	2.1	1116	15	AAW60608
31	185	2.1	2431	13	AAW25138
32	184	2.0	1390	18	AAW21636
33	184	2.0	2237	21	AAW58148
34	180	2.0	501	22	AAW31999
35	178	2.0	34	14	AAW44728
36	178	2.0	501	22	AAW31996
37	177.5	2.0	525	21	AAV91964
38	176.5	2.0	935	21	AAW07570
39	165.5	1.8	255	20	AAW87729
40	163	1.8	1066	21	AAW07561
41	162	1.8	1456	14	AAW34129
42	161.5	1.8	1128	11	AAW05107
43	161	1.8	3266	21	AAW42491
44					
45					

ALIGNMENTS

RESULT 1	AAV31381	standard; Protein; 1698 AA.
ID	AAV31381	
XX	AAV31381	
AC	12-OCT-1999	(first entry)
XX		
DE	HEV-US1 ORF1 protein.	
XX		
KW	Hepatitis E virus; HEV; binding partner; virus; US-HEV infection;	
KW	vaccine; passive immunisation.	
XX		
OS	Hepatitis E virus.	
XX		
FH	Key	Location/Qualifiers
FT	Misc-difference 174	/label="unknown"
FT	Misc-difference 363	/note="encoded by ATR"
FT	Misc-difference 1131	/label="unknown"
FT	Misc-difference 1088	/note="encoded by RTG"
FT	Misc-difference 1217	/label="unknown"
FT	Misc-difference 1389	/note="unknown"
FT	Misc-difference 1389	/label="unknown"
FT	Misc-difference 1389	/note="encoded by GMC"

Protein encoded by
Encoded by ORF 1 O
Protein encoded by
Protein encoded by
Hepatitis E virus
Protein encoded by
HEV isolate 14401
HEV isolate 2015-1
HEV isolate 212-or
Sequence of Heliot
Rubella virus RA27
Amino acid sequenc
Amino acid sequenc
Infectious rubella
HEV ORF1 peptide (Nudaurelia beta-1i Nudaurelia beta vi Amino acid sequenc Amino acid sequenc Rupesitris stem pit Tobamovirus replic SFV4 non-structura Grapevine leafroll b GRFav-3 polypeptei Amino acid sequenc HEV ORF1 peptide (Amino acid sequenc PMNAV-1 RNA-depend Protein encoded by Rupesitris stem pit Protein encoded by PVX replicase. Po Sequence encoded b Human ORFX ORF2255

XX MO9919732-A1.
 PN 22-APR-1999.
 XX 15-OCT-1998; 98WO-US21941.
 PF 15-OCT-1997; 97US-0061199.
 XX (ABBO) ABBOTT LAB.
 XX Dawson GJ, Desai SM, Erker JC, Mushahwar IK, Schlauder GC;
 DR WPI: 1999-288017/24.
 DR N-PSDB: AA200195.
 XX
 PT Detection of United States Isolates of hepatitis E virus
 PS Claim 6; Page 175-180; 260pp; English.
 XX
 CC The invention provides a method for detecting a US (sub)type hepatitis E
 CC virus (US-HEV), or its naturally occurring variants in a sample by
 CC treatment with a binding partner specific for a marker of the virus, and
 CC then detecting any complex formed. The method is used to diagnose
 CC infection with US-HEV. Polypeptides from US-HEV, antibodies specific for
 CC open reading frames (ORF) in US-HEV and host cells expressing these ORFs
 CC are useful in vaccines or for passive immunisation. The polypeptides are
 CC also used to raise specific antibodies (useful as immunoassay reagents).
 CC Fragments of nucleic acid from US-HEV are useful as primers and probes in
 CC usual hybridisation and amplification assays for detecting infection. The
 CC present sequence represents a HEV-US1 ORF1 protein.
 XX
 SQ Sequence 1698 AA;

Query Match 99.9%; Score 9001; DB 20; Length 1698;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1698; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PGITTAIEQALAAANSAANANVVRPELRSVOTELLINLMOPROLVFERPEVLMNHPIOR 60
 DB 1 pgitatgaalaaansalanavvrpfisrvtellinlmprqivfpevilmnhpiqr 60
 QY 61 VTHNEBOYCARAGRCLEVGAPHSINDPNVLRHCFRPVGRDVOYMSAPTRGPAN 120
 DB 61 vthneleqycaragrclevgahpsindpnvlrhcflrpvgrdvqwyasaptrgpaan 120
 QY 121 GRSRLRGIPADRTYCPFGFSRCAFAAFTGYALSLHDMPADVAEAMARRGXTRLYA 180
 DB 121 grrsrlrgipadrtycpfgfsrcafaaftgyalslhdmpadvaeamarrgxttrlyaa 180
 QY 181 LHLPEEVLPPGTYHTTSYLIIHGDRAVVTVEGDTSAAGYNHDSVILIRAMIRTTKIVGDH 240
 DB 181 lhlpeevllppgtyhttsyllihgdgravvtvegdtsaagynhdsvillramirttkivgh 240
 QY 241 PLVIERVRACGHEVULLTAAPESPMPYVPRSTEVYVRSIFGPGSPSLFPSACSTK 300
 DB 241 plviervracghevulltaapespmpypvrstevyvrslfpggspslfpsacstk 300
 QY 301 SEFHVPPVHIMDRMLFGATLDDOAFCCSRMLTYLRGISYKTVGALVANEGMNASEDL 360
 DB 301 sfhvpvphimdrmlfgatlddoafccsrmltylrgisylvkvgalvanegmnasedl 360
 QY 361 TAXITAAVLTICHOYRLFTQATISKMRRLGVEHAOKFIRLXSMLEPKSGRIYIGRQ 420
 DB 361 taxitaaavlitchoryrlftqatiskmrrlgvehaokfirlxswlfpksgriyigrq 420
 QY 421 FYAOCRRMLISAGFHLDPRLVFDSEVPCCRTFLKVKAGKFCFCFMRMLQOECTCELEPAE 480
 DB 421 fyaocrrmlisagfhlprlvfdesvpccrtflkvvagkfccfcmrmlqoectcelfepe 480
 QY 481 GLVGHGHDNEVYEGSEVDPAEPALDVSCTYAVGHQLEALYRLANVPDITAAASRLT 540
 DB 481 glvghghdnevyegeevdpaealdvsgtyavghqlealylrlanvpditaasrlt 540

DB 481 glvghghdnevyegeevdpaealdvsgtyavghqlealylrlanvpditaasrlt 540
 QY 541 ATVELVASPDRLCEKRTVLGNKTFRTTVYDGAHLLEANGPEQOYLSFDSARQSMGAGSHSLT 600
 DB 541 atvelvaspdrlcekrvtlgnktrtvtvydgaahlleangpeqoylsfdsarqsmgagshslt 600
 QY 601 YELTPAGLOVRISNSGLDCTAATPPPGAPSAAPGEVAAFCALYRYNRTQSHSLTGLIM 660
 DB 601 yeltpaglvrisnsgldctaattpppgapsaapgevaafcallyrynrtqshsltglw 660
 QY 661 LHPGGLGTFPPSPGHIMESANPFCGECTLYTKRMSTSGFSDDSPPRAAPAMAAATPG 720
 DB 661 lhpgllglfppspghimesanpfcgectlytkrmstsgfsdpsppraapamaatpg 720
 QY 721 LPHSTPVSIDIWLPPESEFQVDAAPVPADPAGLPGPVLTLPPEPPVVKPSLPPPS 780
 DB 721 lphstpvsvdiwlppesefqvdaappvpadpaglpgpvltpppppvvkpslppps 780
 QY 781 RNRRLLYTPDGAKYAGSLFESDCDMLVNASNPGHRPGGLCHAFYORFPAFTPTETI 840
 DB 781 rnrillytpdgakayagslfesdcawlvnasnpghrpggichafyorfpaftpteti 840
 QY 841 MREGIAAYITPRPIIHAAPDYRVEONPKRLEAAVRETCRSRGTAAAYPLLSGTYQVPV 900
 DB 841 mregiaayitprpiihavapyrveonpkrlleaavretcsrgrtaaypllsqylyqvpv 900
 QY 901 SLSPFAMERNRHPGDELYLTPEPANWFENKPAOPVLTITEDTARTANALEIDAATEVG 960
 DB 901 slsfawernhrpgdeyltpepanwfeankpaoqvltitedtartaanaleidaatevg 960
 QY 961 RACACGTTISPGIVHOFNAGVSGSKSRNIOGDVDVYVPPRELNSMRKRGFAFTTH 1020
 DB 961 racacgttispgivhofnagvsgsksrniogddvvyvpprelnsmrkrgfaftth 1020
 QY 1021 TAARVTIGRRVYIDEAPSLPHLLLLHMQRASVHLLGDPNOIPIAIDFHAGIYVAIRPE 1080
 DB 1021 taarvtigrvyyideapslphllllmqrasvhlldgdpnoipiaidfahagilvaipre 1080
 QY 1081 IAPTSMWXYVTHRCPADVCELRIGAYPKIOTTSRVLSLEFWNPAPAIQKLVXTQAAKANP 1140
 DB 1081 iaptsmwxyvthrcpadvcelrigaypkioettsrvlslefwnpapaiqklvxtqaakaanp 1140
 QY 1141 GAITVHEOGATFTETTTIATADAGLIQSSAHAAIVALTRTEKCVIILDAGLLREVG 1200
 DB 1141 gaitvheogatftetttiatadagliqssahaaivaltrtekcvilidagllrevgi 1200
 QY 1201 SDVINNFELAGCEYGHSPSYIPRGNPNQDGTIOAPFPCOISAYHOLAELEGHRPAP 1260
 DB 1201 sdvinnfelagceyghspsyiprgnnpndgtioapfpcoisayholaeleghrapap 1260
 QY 1261 VAAVLPPCEBELBQGLLYMPQELTVSDVLFELTDIVHCMAAPSORKAVLSTLVGRYGR 1320
 DB 1261 vaavlppebelbqgllympqeltvsvdlfeltdivhcmappsrkavlstlvgrgr 1320
 QY 1321 RRTKLEAAHSDVRESLARTIPRTIGVQATTCGLYELVEMVNRKQDGSVLELDCNRY 1380
 DB 1321 rrtkleaahsdvreslartiprtigvqatcglylevnmvnrkqdgsvleldcncry 1380
 QY 1381 SRTEFOKKCNKFTTGETAHAGKVGOGISAMSKTCALGPMFRAIERKIIILAPNIFY 1440
 DB 1381 srteffokkcnkfttgetahagkvogogisamskctalgpmfriaerkiilapnify 1440
 QY 1441 GDAYEESVFAAASGAGSCMVENFSEEDSYQNNFSLGLECVNMECGMOMLRLYHL 1500
 DB 1441 gdayeesvfaaasgagscmvenfseedsyqnnfslglecvnmeecgmpomlrllyhl 1500
 QY 1501 VRSANITLAPKESLKGPMKKHSGEPGTLIMNTVMMAITAHCEYRDERVAAFKGDDSYV 1560
 DB 1501 vrsaniltapkeslkgpmkshgpgtllmntvmmaitaahceyrdervaaafkgddsvv 1560
 QY 1561 LCSDRQSRNAALALGAGGKIKVYRPIGLYAGVVAAPGLTLDVYVFAARLSEKMWG 1620
 DB 1561 lcsdyrqsrnaaalalgagkiklvyrpigliyagvvapgltdlvvyfaarlsekmwg 1620

1621 PGERAEOLRLAVCDFRLGLTNVAQCVDSRVGSPBELVHNILGMLQTIADGKAHET 1680
1621 PGERAEGLRLAVCDFRLGLTNVAQCVDSRVGSPBELVHNILGMLQTIADGKAHET 1680
1681 ERIKPYLDLNTSIQORVE 1698
1681 etikpyldlntsiqrve 1698

RESULT 2
AAM93408
ID AAM93408 standard; Protein; 1708 AA.
XX
AC AAM93408;
XX
DT 11-JUN-1999 (first entry)
XX
DE Swine HEV ORF 1 protein.
XX
KM Swine hepatitis E virus; HEV; cross-reaction; antibody; human; therapy;
KM vaccine; immunise; infection; detection; diagnosis; prevention.
OS Hepatitis E virus.
XX
- PN MO9904029-A2.
XX
PD 28-JAN-1999.
XX
* XX 17-JUL-1998; 98WO-US14665.
XX
PR 18-JUL-1997; 97US-0053069.
XX
FA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Emerson SU, Meng X, Purcell RH;
XX
DR WPI: 1999-132270/11.
XX
PT New isolated swine hepatitis E virus - used to develop products for
PT the diagnosis, prevention and treatment of hepatitis E virus
PT infection in mammals, particularly humans
PS Example 1; Fig 6D-J; 70pp; English.
XX
XX This invention describes a swine hepatitis E virus (HEV) and its natural
CC mutants which are capable of cross-reacting with antibodies reactive
CC with a human HEV strain or natural mutants. The HEV and the proteins
CC can be used in vaccines for immunising against HEV infection. The swine
CC HEV can be used in humans to prevent possible infection by human HEV. The
CC swine HEV can also be used as a therapeutic treatment for infection by
CC other strains of HEV. The swine HEV can also be used for the production
CC of antibodies which can be used in therapy, detection and diagnosis. The
CC products can also be used for determining the susceptibility of cells or
CC organs to infection with swine HEV. The swine HEV is particularly useful
CC for the development of agents for the prevention, treatment and detection
CC of human HEV because it is not a human virus and thus can be handled both
CC experimentally and clinically without fear of severe infection and/or
CC contamination.
CC
SQ Sequence 1708 AA;

Query Match 97.8%; Score 8812.5; DB 20; Length 1708;
Best Local Similarity 97.8%; Pred. No. 0;
Matches 1661; Conservative 7; Mismatches 30; Indels 1; Gaps 1;

1 PGTTTAEQAALAAANANVVRPFSLSRQTELLINLMQPROLVFPEVIMNPIQR 60
10 pgtttaeqaalaaanavvrpfslsrqteellinlmqprolvfpevimnpiqr 69
61 VTHNELEQYCARAGRCLEVGAPRISINDPNVNLHRCFLRPVGQVDQRMYSAPTRGPAAN 120
VTHNELEQYCARAGRCLEVGAPRISINDPNVNLHRCFLRPVGQVDQRMYSAPTRGPAAN 120

70 aihneleqyrcraragcclevgauprfindbnpvlnhrcflrpvgqvdqrmysaptrgpaan 129
121 CRRSALRGLEPPADRTTCFDFSRCAFAETGVALYSLHDIIMPADVAPAMARHCXTRLYAA 180
130 crrsalrgleppadrttcfdgfsrcafaetgvallyshdiimpadvaamarmhgmtrlyaa 189
181 LHLPPFVLLPFGVYHTTSYLLIHDGRAVYTESDTSAGYNNHDSILRAMTRTKTYGDH 240
190 lhlppfvllpfgvhttsyllihdgravytesdtsagynnhdsilramtrtktygdh 249
241 PLVIERVRAIGCHFVLLTAAPESPMPVYVPRSTEVYRSTFGPGSPSLFPSACSTK 300
250 plviervraigchfvlltaapepmpvyvprstevyrstfpgpgpspslfpasactk 309
301 STFHAAPVHIWDRIMLFATLDDQAFCCSRLMYLRGISYKVTYGVALVANEWNASDAL 360
310 stfhaapvhiwdrimlfgatlddqaafccsrlmylrgisykvtvgvalvanegwnasdal 369
361 TAXITAAVLTICHQRYLRTQAGISKGMRLGVEHAOKRTITLWSLMEFKSGRDYIPGRLO 420
370 taxitaaavlthcqrlylrtqagiskgmrlgvehaokrtitlwslmefksgrdyipgrlq 429
421 FYACRRRLSAGFHLDRVLVFDSEVPCRCRTFLKTVAGKFCCEPMRLGCECTCLEPAE 480
430 fyacrrrlsagfhlldrvlvfdsevpccrtflktvagkfccetmrvlqgeccclepae 489
481 GLVGDHGHNEAVEGSEVDPAEPALHDVSGTYAVHGQLEALRYALNVPODIAARASRLT 540
490 glvgdghghneavegsevdpaepalhdvsgtyavhghqlealryalnvpoidiaarasrlt 549
541 ATVEIVASPDRLFCRTYLGKTRRTTYVDGAHLEANPEQYVLSFDSRSGMAGSISLT 600
550 atveivaspdrlfcrtylgktrrttyvdgahleanpeqyvlsfdsrsgmagsislt 609
601 YELTPAGLQYRISNGLDLCATFPFGGAPSAAGENVAFCSALYRYRRTFORHSLTGLW 660
610 yeltpaglqyrisngldlcatfppggapsaagevafcsalryrrtfgrhsltgglw 669
661 LHPBGLGIFPPSPGHIWESANPFCGEGTLYRTWSTSGFSSDSFPEEAAPAMAATPG 720
670 lhpglglgifppspghiwesanpfcgegltlyrtwstsgfssdsfpeeaapamaatpg 729
721 LPHSTPVSQIWLPPPESEFOYDAAPVPAPAPAGIPGVVLT-PPPPPVMKPSLPP 779
730 lphstpvswlpppsesfoydaavpappapagipgvvlt-ppppvmkpslpp 789
780 SNRRRLTYTPDDAKYVAGSLFESDCDMLVNASNPGRRPGGICHAAYOFPEAFTYEF 839
790 snrrrltytpddakvyagslfesdcmlvnaasnpgrrpggichayofpeaftyef 849
840 IMREGLAAYTLTPRPIIHAAVPDYRVEQNPRLAEAVRETCSRRTAAYPLGSGIYQVP 899
850 imreglaaytltpripiahavpdyrveqnpkrlaeavretcsrrtaaypllgsgiyqvp 909
900 VLSFPAWENHNRPGDELTYTEPAANMEFANKRAQVLTTEPTARANALALEIDATEV 959
910 vlsfpawenhnrpgdeltytepaanmefankraqvltteptaranalaleidatev 969
960 GRACACCTISPGIYVHOFTGVGSGKRSITQCGDVVVVVVPRRELNSRRRGFAFTP 1019
970 gracacctispgiyvhofgvsgkrsitqcgdvvvvvprrelnsrrrgfaftp 1029
1020 HTAARVYTGRRVYIDEAPSLPPLLHLMQRASSVHLGDPNOIPALDFEHAELVPAIRP 1079
1030 htaarvytgrvyyideapslppllhmqraassvhlgdpnoipaldfehaelvpaairp 1089
1080 ELAPTSWMMXYTHRCPADVCLIRGAYPKIOTTSRVLSLWNEPAIGOKLYVIOAKAAN 1139
1090 elaptswmxythrcpadvclirgaypkiottsrvlslwnepaigoklyvfioakaan 1149
1140 PGALIVHEAGAFTEFTTITATADARGLDSSAHMIVALTTRHEKCVILDAGGLREVG 1199
1150 pgalivheagaftefttitaadarglldssahmivalttrhekcvildagglirevg 1209

QY	1200	ISDVIYNNFELAGEVXCHHPSPVTPGPNPNQNLCTLOAPPEPCOSIAVHOLAEELGHBA	1259
Dd	1210	ISDVIYNNFELAGEVXCHHPSPVTPGPNPNQNLCTLOAPPEPCOSIAVHOLAEELGHBA	1269
QY	1260	PVAAPVLPCCPELEGGLLYMPQOELTVSDSVLYFELTDIVHCRMAAESORKAVLSTLVGRYG	1319
Dd	1270	pvaavlpccpeleqllymppqeltvdsavlyfeltdivhcrmaapsgqkavlstlvgryg	1329
QY	1320	RRTKLTYEAHSDVRESIARITPTIGVQATTCGLTYELVYAMERKODSDAVLEDLCCNRD	1379
Dd	1330	rrtklyeahsdrvresiaritptigvqatcgltyelvymerekodsdavleldlccnr	1389
QY	1380	VSRTTFPOKXCNKFTTGETTETIAHGKVCOGISAMSRTFCALFCGFWFAIEKEIILALPNNIF	1439
Dd	1390	vsrttfpdkcnkfttgettetiahgkvogisamsrtfcalfcgfwfaikeiilalppnlf	1449
QY	1440	YGDAYEESVFPAAVVSAGSGSCMFENDFSEPDSTQNNFISGLECYVMEBCGPMOILRLTX	1499
Dd	1450	ygdayeessvfpaaavvsagscmvfendfsepdstqnnfisgleyvmeecgpmviltlrx	1509
QY	1500	LVRSAWILLOAPKESIKCFMKHKGEGGTLIMNTVMNMALIALICYEFPRFVAAEFKGDSDV	1559
Dd	1510	lvrsawilgapkeslkgfwkhhsgsegtllmntvnmalmialicyeftrdrvaefkgsdv	1569
QY	1560	VLCSDYRQSRMAALLINGCGLIKLVDRPRLGLYAGVYVAPGICLTPDVRYPFRAGRISSENM	1619
Dd	1570	vlsdyrqsrmaalllagcgliklvdyrplglyagvyvapgicltpdvyrpfragrisesnm	1629
QY	1620	GPGERAEOLLELAWCDELRLGLTVNAQCVDVVSRRYGVSPGLVHNLICMLQTIADKANH	1679
Dd	1630	gpgeraeqlrlawcdelrlgltnvaqcvdvvsrrygvspglvnhlimglqladgknhf	1689
QY	1680	TEITIKPVLDLTNSITIQRYE	1698
Dd	1690	teitlkpvldltnsitlqrye	1708

-RESULT	3
AAV31384	
"ID	AAV31384 standard; Protein; 1708 AA

AC AAY31384;

DT 12-OCT-1999 (first entry)

HEV-US2 ORF1 protein.

KW Hepatitis E virus; HEV; binding partner; virus; US-HEV infection;

KW vaccine; passive immunisation.

OS Hepatitis E virus.

FH	key	Location/Qualifiers

FT /label= unknown

FT	Misc-difference	331
----	-----------------	-----

FT	encoded by RAC"	/note=

FT	/label= unknown
----	-----------------

FT Misc-difference 448 / more - encoded by nmc

E1	/label= unknown
E1	/note= "encoded by RTG"

```

FI MISCELLANEOUS 034
FI /label = unknown
FI

```

FT	/note= "encoded by CCR"
FT	Misc-difference 646

```
ET /label= unknown
ET /note= "encoded by CCS"
```

FT MSc-difference 811

FT	/label= unknown
FT	/note= "encoded by TKR"
FT	Misc-difference 1533
FT	/label= unknown
FT	/note= "encoded by TAY"
FT	Misc-difference 1578
FT	/label= unknown
FT	/note= "encoded by RGC"
FT	Misc-difference 1691
FT	/label= unknown
FT	/note= "encoded by GAR"
PN	W09919732-A1.
PD	22-APR-1999.
PE	15-OCT-1998; 98MO-US21941.
PR	15-OCT-1997; 97US-0061199.
PA	(ABBO) ABBOTT LAB.
PI	Dawson GJ, Desai SM, Eiker JC, Mushahwar IK, Schlauder GG;
DR	WPI: 1999-288017/24..
DR	N-PSDB: AAZ00267.
XX	Detection of United States Isolates of hepatitis E virus
XX	Claim 6; Page 216-221. 260pp; English.
XX	The invention provides a method for detecting a US (sub)type hepatitis E
XX	virus (US-HEV), or its naturally occurring variants in a sample by
XX	treatment with a binding partner specific for a marker of the virus, and
XX	then detecting any complex formed. The method is used to diagnose
XX	infection with US-HEV. Polypeptides from US-HEV, antibodies specific for
XX	open reading frames (ORF) in US-HEV and host cells expressing these ORFs
XX	are useful in vaccines or for passive immunisation. The polypeptides are
XX	also used to raise specific antibodies (useful as immunoassay reagents).
XX	Fragments of nucleic acid from US-HEV are useful as primers and probes in
XX	usual hybridisation and amplification assays for detecting infection. The
XX	present sequence represents a HEV-US2 ORF1 protein.
XX	Sequence 1708 AA;
XX	50

Query Match	97.1%;	Score 8753.5;	DB 20;	Length 1708;
Best Local Similarity	97.2%;	Pred. No. 0;		
Matches 1651; Conservative	8;	Mismatches 39;	Indels 1;	Gaps 1

QY 1 PGITTAIEQAAALAAANSALANAVVRPFLSRVQTEILINIMQPRQLVFRPEVLWNHPQR 60

Db 10 pgittaalegaalaansalanavvrpflsrvgtellnlmqprqlvfrpevlwnhpqr 69

QY 61 VINHELECYCRARAGRCLEVGHAHPRSINDNPVLHRCFLRPVGGRDVQRWYSAPTGPAN 120

Db 70 vihneleqycrararqrclevqahprrsindupnv1hrcflrpvqrdvqrwysapttrqpaan 129

ov 121 CRSALRGLPADRTYCFDEGSRCAFAETGVALYSLHDLMPADVAEAMARHGTRLAA 180

Dp 130 crsralrqlppvdrtycfdfsrcafaaetqvalyslhdlpwadvaeamarhamtrlyaa 189

181 T.H.PREV.T.PRGTVHTTSY.T.THDGBAVTVYEGDTSAGYNHDVSTI.BAWTBTKIVGDH 240

```

190 |h|nev|]p|atv|ht|sv|]i|h|a|n|r|a|v|t|v|e|q|d|s|a|v|n|h|d|v|s|i|]r|a|w|i|r|t|t|k|i|v|d|h| 249

```

00 341 RYTERBVAICCHUW I I TADBDSDMBVBPSTEFVVPSTFCBGCSPSLFBSACSTK 300

```

350  njvjerura jicbfvllltaapensmuvnrstaurvreiffoaaensl fnsacstk 309

```

361 СЕРУАИДИТИДИ ИР ПОАЕР ДООА ПОООССТ ИВУР ПОТОВИИИКАЛ ВАНЕСИНАСЕДАТ 360

[illegible]

```

QY 361 TAXITAAVLTICHOVRLTQATISKGMRLGVEHQAQKITRLYSWLEFKSGROVIRPQO 420
  |||||||
Db 370 favilaaylticqyrlrqaalskgmrlevehaqkftrlyswlfxksgrdyipqrlyq 429
QY 421 FYAOCRRMLASAGFHLDPRLVFEDESVPCHRCRTFLKKVAGKFCCEFMWLOCECFLEPPE 480
  |||||||
Db 430 fyaqctrwlsagfhlkprxlrvfdesvpcrtcfllkvaqkfccfmwlyqectcflepe 489
QY 481 GLVGDHGDNEAYEGSEVDPAEPALDVSGTYAVHGHOLEALYRALNVPDIAARASRLT 540
  |||||||
Db 490 glvgdghdneayegsevdpaeapahldvsgtyavbhqlealyrainlvphdiaarasrlt 549
QY 541 AATVELVASDRLECRFTVGNKTRFTVVGCAHLEANGPPQYVLSFPAASQSMAGSHSLT 600
  |||||||
Db 550 aatvelvasdrlectrvgnktrftlvvgahleangpseevylsfdaasqsmagshslt 609
QY 601 YELTPAGLQVRISSNGLDQTATFPFGAPSAAPGEVAAFCSALYRNRTFORSLTGLW 660
  |||||||
Db 610 yeltpaglvqkissngldctatfpxgapsaapgevaxacsalyrnrftqrhsltgylw 669
QY 661 LHPEGLLGIFPPSPGHIWESANPECGEGLYTRWSTSGFSDFSPPEAAPAMAAATPG 720
  |||||||
Db 670 lhpegllygifppspghiwesanpfcgegllytrwstsgfsdfsppeaapasaapq 729
QY 721 LPHSTPPVSDIWLVPSPSEFOVDAPVPAPDPAGLPGPVLT - PPPPPVHKPSITPP 779
  |||||||
Db 730 lphstppvsdlwlvppspseeshvdaasvpspepagltsplyltppppppvkpsitpp 789
QY 780 SHNRRLLYTPDGAKYVASLFEESDCMLVNASNPGHRCGGICHAFFYRFPAPFPTPE 839
  |||||||
Db 790 ptnrlllytpdgakvyaslfesdcclvnasnpghrpggqichafyrfpeafstef 849
QY 840 IMREGIAATLTPRPITIAVADYREONPKRLAAAYRETCSRGTAAVPLLGSGTYQVP 899
  |||||||
Db 850 imreglaaytlprplihavadyreognpkrlaayretcsrgtaavpllgsgtyqvp 909
QY 900 VELSPFAMRNHRPGDELYTERPAAMFEANKPAQPVLTTEPTATNMLEIDATEY 959
  |||||||
Db 910 velstfdawernhrpgdelylepaawfeankpaqpalitedatnmlaleidatey 969
QY 960 GRACAGCTTSPGIVHKOFAVPGSGKSRSIOGDVVVVPRRELRNSMRRRGFAATP 1019
  |||||||
Db 970 gracagcttspgivyhkofovpgsgksrsiqgdvdvvvprrelnsmrrrgfaatp 1029
QY 1020 HPAARVYTIGRRVYIDEPALPHLLLLHMQRASVHLLGDPNQIPAIIDEHAGLVPAIRP 1079
  |||||||
Db 1030 hpaarvlytigrvvyideapllphllllhmqrassvhlldgdpnqipaidehaglvpaairp 1089
QY 1080 ELAPTSMWKVTYHRCPADVCELIRGAPKTIQTSRYLRSIFWNEPAIGOKLYTOAKKAN 1139
  |||||||
Db 1090 elaptsmwvthtrcpadvcelirgapkltqtsrylrslfwmepaigoklvtfqaakaan 1149
QY 1140 PGATVHHEOGATFTEETTLIATADARGLIQSSRAHAIVALTRHTECVLLIDAPGLREVG 1199
  |||||||
Db 1150 pgaivheogatftecetliatadargliqssrahaivaltrhteccvllidagplirevg 1209
QY 1200 ISDVIVNNEFLAGGEVGHHRPSVIRGNPDONLGLQAPPPSCQISAYQOLAELGHRPA 1259
  |||||||
Db 1210 isdvivnneflaggevgghrpsvirgnpdnlgltqapppesqisayqleelghrpa 1269
QY 1260 PYAAVILPPELEOGILLMPDELIVSDSVLVELDIIHCRMAAPSORKAVSTLVGRYG 1319
  |||||||
Db 1270 pyaavilppeleogillympdelivsdsvlveltdiivnormaapqsrkavstlvgyrg 1329
QY 1320 RRTKLYEAHSDVRESLAFIPTIGVQATCELYELVAMVEKGGDSAVLELDICND 1379
  |||||||
Db 1330 rrtklyeaansdvreslafiptigvratceleyelvamvekqgdsgavleldicnd 1389
QY 1380 VSRITFEFOXKCKFTTGETIAHGKVGQGISAMSKFCALEFQWFAIEEIIALLPNTIF 1439
  |||||||
Db 1390 vsritffgdkcnkfttgetiangkvqgisawsktcalfgwfaieeiiallpntif 1449

```

```

QY 1440 YGDVEESVFAAASGAGSCMFENDFSPEDSTQNNPSLGLCEVWMECGMPWLIRYH 1499
  |||||||
Db 1450 ygdvaeesvfaaavsgagscmfendfsefdestqnnpslglecvmeeccmpwllrlyh 1509
QY 1500 IYRSANILQAPKESLKGFMKHHSEPGTLLMNTVMNNAITAHCEYERDFVAAFKGDSDV 1559
  |||||||
Db 1510 lyrsanilqapkeslkgfmkhhsepgtllmnlvmmaitahceyefdfvaafrkgsdsv 1569
QY 1560 VILCSDRQSRNNAALJAGCGKLKLVDRPIGLYAGVVAAGLCTLPDYVFAGRLEKMW 1619
  |||||||
Db 1570 vilcsdyrxrnnaaallagcgllkvlvdrpigliyagvvaapglctlpdyvfaagrlsekw 1629
QY 1620 GPGEPEAEQURLAVCPDLRGLTVAQVYVSRYVGSGLVHNLIGMLQTTADGKAHF 1679
  |||||||
Db 1630 gpgepeaeqlrlavcdfllrgltlnvaqvcvdsrvyvspglvhnligmlqtadgkahl 1689
QY 1680 TETIKPVLDTNSIIORVE 1698
  |||||||
Db 1690 tnikpvlcltntsiigrve 1708

RESULT 4
AAR91813
ID AAR91813 standard; Protein; 1693 AA.
XX
AC AAR91813;
XX
DT 26-NOV-1996 (first entry)
XX
DE Hepatitis E virus strain SAR-55 ORF-1.
XX
KW Hepatitis E virus; HEV; SAR-55 strain; enteric transmission;
KW structural region; antigen; detection; antibody; vaccine;
KW immunisation; infection.
XX
OS Hepatitis E virus.
XX
FH key Location/Qualifiers
FT Misc-difference 1238 /note="corresponding codon CAG"
FT FT
FT Misc-difference 1455..1693
FT FT
FT Misc-difference /note="10 bp nucleic acid sequence TGCTNTTCA
FT FT has to be inserted between nucleotides
FT FT 4390..4391 of AAT27394 before these amino
FT FT acid residues can be decoded"

MO9610580-A2.
XX
PD 11-APR-1996:
XX
XX 03-OCT-1995; 95WO-US13102.
XX PF
XX PR 03-OCT-1994; 94US-0316765.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Emerson SU, Purcell RH, Tsarev SA;
XX WPI: 1996-209320/21.
XX DR N-PSDB; AAT27394.
XX
XX Isolated and purified hepatitis E virus strain SAR-55 DNA - encodes
XX PT antigenic protein useful in diagnosis, prophylaxis and treatment of
XX PT hepatitis E virus infection
XX
XX PS Disclosure; Pages 9-13; 121pp; English.
XX
XX The present sequence is the protein prod. of ORF-1 from the
XX CC hepatitis E virus (HEV) strain SAR-55, which was implicated in an
XX CC enterically transmitted non-A, non-B hepatitis in Pakistan. The
XX CC protein encoded by the structural region of the virus (i.e. ORF-2),
XX CC which is capable of forming HEV like particles, is useful for the
XX CC detection of HEV antibodies (pref. IgG or IgM) in blood, plasma,

```

CC sera, cerebrospinal fluid, tissue, urine or pleural fluid. The
 CC protein, and anti-HEV antibodies generated using the protein, can
 CC also be used in vaccines for immunising an animal against HEV
 CC infection. The protein is identified as a band of greater than
 CC 50 kD following SDS-PAGE of cell lysates of insect cells infected
 CC with a HEV ORF-2 contg. baculovirus, i.e. the claimed recombinant
 CC expression vectors pPIC9-1779, -1780 and -1781.

XX Sequence 1693 AA;

Query Match 82.7%; Score 7456; DB 17; Length 1693;
 Best Local Similarity 81.8%; Pred. No. 0;
 Matches 1400; Conservative 110; Mismatches 160; Indels 42; Gaps 7;

QY 1 PGITTAIDQALAAANSALANAVVRPELSRVOTELLMLMPROLVPEVLMNPIOR 60
 DB 10 PGLTALIEQAALAAANSALANAVVRPFLSHQJELLINLMGRQLVITPEVFMHPIDR 69
 QY 61 VTHNELEQCRARAGRCLEVGARPSINDPNVLRHRCFLRPVGRDQORVYSAPTGPAAN 120
 DB 70 VTHNELEQCRARAGRCLEVGARPSINDPNVLRHRCFLRPVGRDQORVYSAPTGPAAN 129
 QY 121 CRRSALRGLPPADRTYCEDFGSRCAFAETGVALYSLHDLMPADVAEAMARAGXTLYAA 180
 DB 130 CRRSALRGLPPADRTYCEDFGSRCAFAETGVALYSLHDLMPADVAEAMARAGXTLYAA 189
 QY 181 LHLPEVLLPCTYHTTSTYLLHGDRAVYTESGTSAGYNNDVSTLRMTTKIYGH 240
 DB 190 LHLPEVLLPCTYHTTSTYLLHGDRAVYTESGTSAGYNNDVSTLRMTTKIYGH 249
 QY 241 PLVIRVRAGCFHFLLLAARPEPMPVYPVPRSTEVVRSIFGSGSPSFPSCASTK 300
 DB 250 PLVIRVRAGCFHFLLLAARPEPMPVYPVPRSTEVVRSIFGSGSPSFPSCASTK 309
 QY 301 STEFNAVPVHTMDRLMFGATLDDQAPCCSRMLTYLRGISYKVTGVALNBSGWNASEDAL 360
 DB 310 STEFNAVPVHTMDRLMFGATLDDQAPCCSRMLTYLRGISYKVTGVALNBSGWNASEDAL 369
 QY 361 TXHIAAYLTICHOHRLFRQALSKGRMRGLVGHAKFIRLXSMLEKSGRIYIGRQAO 420
 DB 370 TXHIAAYLTICHOHRLFRQALSKGRMRGLVGHAKFIRLXSMLEKSGRIYIGRQAO 429
 QY 421 FYAQCRRMLASGPHLDPRVLVDESVPCRCRTFLKRVAKFCCEFMKMLQOECTCELEPAE 480
 DB 430 FYAQCRRMLASGPHLDPRVLVDESVPCRCRTFLKRVAKFCCEFMKMLQOECTCELEPAE 489
 QY 481 GLVGHGHDNEAYEGSEVDPAEPALHDVSGTYAVHGHQLEALRYKALNVQDIAARASRLT 540
 DB 490 GLVGHGHDNEAYEGSEVDPAEPALHDVSGTYAVHGHQLEALRYKALNVQDIAARASRLT 549
 QY 541 ATVEVLVAPSDRLCEKTVYGNKTFRTTVVDGAHLEANGRPOYLVSPDASQSGASHSLT 600
 DB 550 ATVEVLVAPSDRLCEKTVYGNKTFRTTVVDGAHLEANGRPOYLVSPDASQSGASHSLT 609
 QY 601 YELTPAGLQVIRISSNGLDCTATFPFGGASAPARGEVAACALYRXNRETOHSLTGLIM 660
 DB 610 YELTPAGLQVIRISSNGLDCTATFPFGGASAPARGEVAACALYRXNRETOHSLTGLIM 669
 QY 661 LHPBGLGIFPPSPGCHWESANPFCGEGTLTTRTWS-----TSGFSDFSPPE 709
 DB 670 LHPBGLGIFPPSPGCHWESANPFCGEGTLTTRTWS-----TSGFSDFSPPE 724
 QY 710 AAAPMAATPGLPSTPVSDIWLPRPSEERQVDAAPV-PPAPPA-GLPAPVULTPPP 767
 DB 725 AAAPMAATPGLPSTPVSDIWLPRPSEERQVDAAPV-PPAPPA-GLPAPVULTPPP 761
 QY 768 PPPVHKPSIP-PPSRNRLLYTPDGAKYVAGSLFESDCDMLVNSNPGHRGAGGICHAFF 826
 DB 762 PPPVHKPSIP-PPSRNRLLYTPDGAKYVAGSLFESDCDMLVNSNPGHRGAGGICHAFF 821
 QY 827 YORPEAFYPTPEFIMREGIAATYTLPRPIIAVADRYEONPKLEAAVRETCSRKGA 886
 DB 827 YORPEAFYPTPEFIMREGIAATYTLPRPIIAVADRYEONPKLEAAVRETCSRKGA 886

DB 822 YGRYPASIFAASfvmrdgaaytlfprihhavapdyrlenhpkrlaaeyretcsrlgta 881
 QY 887 AYPILGSGIYQVPVSLSPDAMERNRPGDELTLTEPAAMWFAANPAQVULTTEEDTART 946
 DB 882 AYPILGSGIYQVPVSLSPDAMERNRPGDELTLTEPAAMWFAANPAQVULTTEEDTART 941
 QY 947 ANLALEIDATVETGRACAGCTTSPGIVHGYFAGVPGSGSKSSIOGDVVVVVPTRELR 1006
 DB 942 ANLALEIDATVETGRACAGCTTSPGIVHGYFAGVPGSGSKSSIOGDVVVVVPTRELR 1001
 QY 1007 NSWRRRGFAAPTPHTAARTTIGRRVVIDEAPSLPHLLILHMQRASSVHLLDPPMOIPAI 1066
 DB 1002 NSWRRRGFAAPTPHTAARTTIGRRVVIDEAPSLPHLLILHMQRASSVHLLDPPMOIPAI 1061
 QY 1067 DEHAGLVPAIRPELAPTSMMWVTHRCPADVDELRGVAPKIQTMSRVLSLFWMPEPAIG 1126
 DB 1062 DEHAGLVPAIRPELAPTSMMWVTHRCPADVDELRGVAPKIQTMSRVLSLFWMPEPAIG 1121
 QY 1127 OKLVYTOAKAANPAATVHEAQAFTTEPTTIAATADAGLIQSSRAHAIVALTHTTEKC 1186
 DB 1122 OKLVYTOAKAANPAATVHEAQAFTTEPTTIAATADAGLIQSSRAHAIVALTHTTEKC 1181
 QY 1187 VILDAAGLIREVIGSDVIVNPFLLAGGEVGHKRPVYIPRGNDQNLGLQAFPPSCQISA 1246
 DB 1182 VILDAAGLIREVIGSDVIVNPFLLAGGEVGHKRPVYIPRGNDQNLGLQAFPPSCQISA 1241
 QY 1247 YHQLAEELGHRPAPVAAVPPCPBELGGLVMPQELTVSDVLEFELMDIVHCRMAAPSO 1306
 DB 1242 YHQLAEELGHRPAPVAAVPPCPBELGGLVMPQELTVSDVLEFELMDIVHCRMAAPSO 1301
 QY 1307 RRAVSLTLVGRYGRRTKLYLEAASDVRESLAFPIITIGPVQATTCELYELVEMAYEKGD 1366
 DB 1302 RRAVSLTLVGRYGRRTKLYLEAASDVRESLAFPIITIGPVQATTCELYELVEMAYEKGD 1361
 QY 1367 GSAVLELDICNDVSRITFPQKXCKNFTTGETIANGKGGQGISANSKFTCALFGWFRAI 1426
 DB 1362 GSAVLELDICNDVSRITFPQKXCKNFTTGETIANGKGGQGISANSKFTCALFGWFRAI 1421
 QY 1427 EKEIITALPPNIFYGDVDAEESVFAAASGAGSCWFEENFSEFDSQNNFSLGLBCVYME 1486
 DB 1422 EKEIITALPPNIFYGDVDAEESVFAAASGAGSCWFEENFSEFDSQNNFSLGLBCVYME 1481
 QY 1487 ECGMPQWLIRLYHLVRSAMITLQAPKESLKGFMKKHSGEPGLLMTNMNAILAHCEYFR 1546
 DB 1482 ECGMPQWLIRLYHLVRSAMITLQAPKESLKGFMKKHSGEPGLLMTNMNAILAHCEYFR 1541
 QY 1547 DEFVAAPFGDSDVIVLCSQVROSNAALILAGGLKLVXDRPTIGYAGVVPARGLGLPD 1606
 DB 1542 DEFVAAPFGDSDVIVLCSQVROSNAALILAGGLKLVXDRPTIGYAGVVPARGLGLPD 1601
 QY 1607 VVRPAGRLSEKMMWGPPEKAEQRLAVCDFLGLTNVAVQVDSRVYGVSPGELVHNL 1666
 DB 1602 VVRPAGRLSEKMMWGPPEKAEQRLAVCDFLGLTNVAVQVDSRVYGVSPGELVHNL 1661
 QY 1667 GMLQTIADGKAHFTETIKPVLDLINSIIQRYE 1698
 DB 1662 GMLQTIADGKAHFTETIKPVLDLINSIIQRYE 1693

RESULT 5

AAW80196 standard; Protein: 1693 AA.

AAW80196;

23-DEC-1998 (first entry)

Protein encoded by ORF1 of ET-NANB (HEV) Burma strain DNA sequence.

Enterically transmitted non A non B hepatitis virus; ET-NANB;

Hepatitis E virus; HEV; Burma HEV isolate; vaccine;

diagnostic probe.

OS Non A non B Hepatitis virus.
XX
PN US5824649-A.
XX
PD 20-OCT-1998.
XX
PF 07-JUN-1995; 950S-0475807.
XX
PR 25-JUL-1994; 94US-0279823.
PR 17-JUN-1988; 88US-0208997.
PR 11-APR-1989; 89US-0336672.
PR 16-JUN-1989; 89US-0367486.
PR 13-OCT-1989; 89US-0420921.
PR 05-JUL-1990; 90US-0505888.
PR 07-JUN-1995; 95US-0475807.
XX
PA (GENE-) GENELABS TECHNOLOGIES INC.
XX
PI Bradley DM, FTY KE, Krawczynski KZ, Reyes GR, Tam A;
PI Yarough PO;
XX
XX WPI; 1998-582589/49.
DR N-PSDB; AAV66321.
XX
PT Hepatitis E virus proteins - useful for diagnosis or vaccine
PT production the virus
XX
PS Claim 22; Columns 57-66; 47pp; English.
XX
CC AAW80196-98 are encoded by the genome of the Burma strain of
CC enterically transmitted non A non B hepatitis virus (ET-NANB)
CC (hepatitis E virus (HEV)). The specification describes an isolated
CC protein which is specifically immunoreactive with antibodies present
CC in individuals infected with HEV and encoded by a sequence contained
CC in an open reading frame (ORF) of an HEV genome. The genome has a
CC sequence that is more than 70% identical to the ORF1 sequence from
CC Burma HEV isolate. The protein is used as a vaccine and a diagnostic
CC probe for ET-NANB.
XX
SQ Sequence 1693 AA;
XX
Query Match 82.6%; Score 7448; DB 19; Length 1693;
Best Local Similarity 81.8%; Pred. No. 0;
Matches 1396; Conservative 110; Mismatches 169; Indels 32; Gaps 6;
QY 1 PGITTAIEQALAAANSALANAVVVRPFISRVOTELINIMOPROLTFREPEVLNHPQR 60
DB 10 pglttalegaalaanaalanavvrpfishqgtellinlmprqlvfrpevtmhpqr 69
QY 61 VTHNELDYCRARAGCLEYGAHPRSTINDPNVLRHCFLRPVGRDVORWYSAPTGCPAN 120
DB 70 vthneldycraragcleygahprstindpnvhrchflrpygrdvgrwytaptrgaan 129
QY 121 CRRSALRGULPPARTYCOFSESCAFEAERGVALSILHDMPADVAEAMRHGCTRXAA 180
DB 130 crrsalrgulppartycodfsgcnfpaetcgialyslhmsspdvaeamrthgmrtlyaa 189
QY 181 LHLPEVLPPGTYHTTSYLLIHGDRAVVTYEGDSAGYNDHVSIRAMIRTTKITYGDH 240
DB 190 lhlpevlppgtyhttsyllihdgravvttyegdsagynhdvsnlrsirttkvqdh 249
QY 241 PLVIERVRAIGHFVLLTLAAPPSPMPVYVPRSTEVVRSIFPGGSPLEPSACSTK 300
DB 250 plviervraigfhfvllltaapepsmpyvpvrstevyrslfpggtpslfptsctsk 309
QY 301 STEHAPVHIWDRMLFGATLDDOAFCCSRLMYINGISYKVTYVGAALVAMEGNASDAL 360
DB 310 stehavpahiwdrmlfgatlddofccsrlmyingisykvtvgtlvamegnasadal 369
QY 361 TANTTAAYLTICHQRYLRTQAIKSGMRRLGVEHAOKFTITLYLWLFKSGRDYIPGRLO 420
DB 370 tavitaaayltichqrylrtqaiskmgmrllerehaokftitllywlwfksgrdyipgrgle 429

QY 421 FYAQCRRMLISAGFHLDPRVLVPDESVPICRRTFLFKKVAAGKFCCEPMRLAGECTCLEPAE 480
DB 430 fyacqrrmlisagfhldprvlvpdesvpicrtrtflfkvaagkfccepmrlagectclepae 489
QY 481 GLVGDHNDNEAEGSEVDPAEPANHLDVSGTYAVHGHQLEALYRALVPPDIAARASRLT 540
DB 490 glvgdhdndneaegsevdpaepanhlldvsgtyavhghqlealylralvppdiaarasrlt 549
QY 541 ATVELYASPDRLRCRYLVAKKTRRTTYVDGAHLHANGPEOYVUSFDSKOSMGASHSLT 600
DB 550 atvelyaspdrlrcrylvakktrrttyvdgaahlhangpeoyvusfsdskosmgashslt 609
QY 601 YELPLPGLOYRISNGLDCATFPPGAPSAAGEVAFCSALYRYRRTFORHSLTGWL 660
DB 610 yaasagleyryvaagldhavlavaypsrtapevalafsalylrtrreagrsltgnlw 669
QY 661 LHPEGLGIFPPSPGHIMESANPFCGEGLYRTWMS-TSGFSSDFSP-----PEAAPA 714
DB 670 fhpegllgltfaptspgjhwesampfcgsestlyrttsewavsapapdlqfmgseps 729
QY 715 MAATPGLPHSTPVSIDIWVLPSPSEEFQVDAAYV-PPAPDPAGLP-GRVVLTPPPPPVH 772
DB 730 raatpfl-----aatpppappdpppsapaalaepasgatag 766
QY 773 KPSIP-PPSRNRRLTYTPPGAKYVAGSLFESDCDMLVNSNREGRGGGLCHAFYGRF 831
DB 767 apaltqtlarhrrllftypgskvtagstestctwlvnsvndhrppggllchafygrp 826
QY 832 EAFVPTFEIRREGIAATLTPRPRIIHAVAPDYREONPKRLEAAYRTGCRBGAAAPL 891
DB 827 asfdaasfmrtdgaatllprplihavapdyreonpkrlaayrtgcrsrlgtaapll 886
QY 892 GSGIYQVNSLSPDAMERNRPPGDELYLTREPANWFEANRPAOPVLITTEDTANTANLAL 951
DB 887 gtglyqypisrtdawernhrppgdeyltrepawfearnpaopvlittdedartanlal 946
QY 952 EIDAAETEVGRACAGCTTSPGIVHQTFTAGYVGSKSSIOGGVDVYVYVTRERANSWR 1011
DB 947 eldsatvgracagctrvtpgvvqyftagvgsksrslcgadavvvvtrlermawr 1006
QY 1012 RGFAPPTPHTAARTITGRVVIDEAPSLPHLLLLHMQARASSVLLGDPOIPAIDEHA 1071
DB 1007 rgfaatphtaatrtvgrrvvideapslphllllmqaratvhlldgppqipaideha 1066
QY 1072 GLVPALRPBELAPTSWVXVYTHRCPADVCELRGAYPKIQTTSRYLRSLFMNEPAIGQLVX 1131
DB 1067 glvpalrpbelpstswxvylthrcpadvcclrgaypkmqtsrslsfwgepavqqlvf 1126
QY 1132 TQAAKANPGALIVHEAOGATFTETITATADAGLIQSSRAAIVALTIRTEKCYLDA 1191
DB 1127 tqaaknpgalivheogattettitattadagllqssraaivaaltirtekcylda 1186
QY 1192 PGLIREVIGSDVYVNNFFLAGGEVGHXRPSTVPGNDONLGTQAPPPSCQISAYHQIA 1251
DB 1187 pglirevigsdvynnfllaggevgxhrrpstvpgndonlgtqapppscqisayhqia 1246
QY 1252 EELGHRPAPYAAVLPCCPELEOGILYMPQELTVSDSLVVELLDIVICRMAAPSQRKAVL 1311
DB 1247 eelghrpyaavlppeleogillypqltcdsvvtecltvlvncmaapqrkav 1306
QY 1312 STLVGRYGRRTKLYEAASDVRESLARFTPTIGVQATTCBELVELVAMEKGGDSAYL 1371
DB 1307 stlvgrygrrtklyeaasdvreslarftptigvqattcelyelvamekggdsayl 1366
QY 1372 ELIDCNRDVARITPEFOKXCKFTTGERTIANGKVGOGISANSKTPCALFGWFRATEEIL 1431
DB 1367 elidcnrdvaritftgdcxkfttgertiangkvvgogisansktpcalfgwfrateail 1426
QY 1432 ALPLPNIFGDAYDEESVFAAASGAGSCAMFENDFSEFSDTONNFSLGLCQVWMEECGMP 1491
DB 1427 allpqnifgdaydeesvfaaasgagscamfendfsefstqgnfslglcqmecgmp 1486

QY	1492	OMTEFLYHLVNSAWTILQAPKRSIKSGFMKKHSGEGCTLLAMTVNMMAITLACIEPRDRRA	1551
Db	1487	qwlrltlrlyhltsawlllqapkesllrgfwkxhsgpctllwlvnmavtlcnydrdrfqa	1546
QY	1552	AFKGGDSVVLSDSYKQSNNAALITAGCGLKAKVDYRPGLTAYGVVAPGJGTPDDVFRFA	1611
Db	1547	afkgdssvlscseyqspgaavlltagcglikvdfirpjlgyagvvpagjlgaldvvrfa	1606
QY	1612	GRLSKNNKPGCPERAEQURLAVCDPLRGJLVNAQVCVDVYSRYGVSGLVHNLIGMLQTT	1671
Db	1607	grltsknnkpgpceraequrlrslavsdflrkltlvraqcvdvysrygvsrgjlvhnligmqla	1666
QY	1672	IADKAKHFTETKPYLDLTNSIIQRVE	1698
Db	1667	vadgkahftesvkpyldltlnsilcrtve	1693
RESULT	6		
AA	AA	AA	AA
AC	AC	AC	AC
AD	AD	AD	AD
AE	AE	AE	AE
AF	AF	AF	AF
AG	AG	AG	AG
AH	AH	AH	AH
AI	AI	AI	AI
AL	AL	AL	AL
AM	AM	AM	AM
AN	AN	AN	AN
AO	AO	AO	AO
AP	AP	AP	AP
AQ	AQ	AQ	AQ
AR	AR	AR	AR
AS	AS	AS	AS
AT	AT	AT	AT
AW	AW	AW	AW
AX	AX	AX	AX
AY	AY	AY	AY
AZ	AZ	AZ	AZ
BA	BA	BA	BA
BB	BB	BB	BB
BC	BC	BC	BC
BD	BD	BD	BD
BE	BE	BE	BE
BF	BF	BF	BF
BG	BG	BG	BG
BH	BH	BH	BH
BI	BI	BI	BI
BJ	BJ	BJ	BJ
BK	BK	BK	BK
BL	BL	BL	BL
BM	BM	BM	BM
BN	BN	BN	BN
BO	BO	BO	BO
BP	BP	BP	BP
BQ	BQ	BQ	BQ
BR	BR	BR	BR
BS	BS	BS	BS
BT	BT	BT	BT
BU	BU	BU	BU
BV	BV	BV	BV
BW	BW	BW	BW
BX	BX	BX	BX
BY	BY	BY	BY
BZ	BZ	BZ	BZ
CA	CA	CA	CA
CB	CB	CB	CB
CC	CC	CC	CC
CD	CD	CD	CD
CE	CE	CE	CE
CF	CF	CF	CF
CG	CG	CG	CG
CH	CH	CH	CH
CI	CI	CI	CI
CJ	CJ	CJ	CJ
CK	CK	CK	CK
CL	CL	CL	CL
CM	CM	CM	CM
CN	CN	CN	CN
CO	CO	CO	CO
CP	CP	CP	CP
CQ	CQ	CQ	CQ
CR	CR	CR	CR
CS	CS	CS	CS
CT	CT	CT	CT
CU	CU	CU	CU
CV	CV	CV	CV
CW	CW	CW	CW
CX	CX	CX	CX
CY	CY	CY	CY
CZ	CZ	CZ	CZ
DA	DA	DA	DA
DB	DB	DB	DB
DC	DC	DC	DC
DD	DD	DD	DD
DE	DE	DE	DE
DF	DF	DF	DF
DG	DG	DG	DG
DH	DH	DH	DH
DI	DI	DI	DI
DJ	DJ	DJ	DJ
DK	DK	DK	DK
DL	DL	DL	DL
DM	DM	DM	DM
DN	DN	DN	DN
DO	DO	DO	DO
DP	DP	DP	DP
DQ	DQ	DQ	DQ
DR	DR	DR	DR
DS	DS	DS	DS
DT	DT	DT	DT
DU	DU	DU	DU
DV	DV	DV	DV
DW	DW	DW	DW
DX	DX	DX	DX
DY	DY	DY	DY
DZ	DZ	DZ	DZ
EA	EA	EA	EA
EB	EB	EB	EB
EC	EC	EC	EC
ED	ED	ED	ED
EE	EE	EE	EE
EF	EF	EF	EF
EG	EG	EG	EG
EH	EH	EH	EH
EI	EI	EI	EI
EJ	EJ	EJ	EJ
EK	EK	EK	EK
EL	EL	EL	EL
EM	EM	EM	EM
EN	EN	EN	EN
EO	EO	EO	EO
EP	EP	EP	EP
EQ	EQ	EQ	EQ
ER	ER	ER	ER
ES	ES	ES	ES
ET	ET	ET	ET
EU	EU	EU	EU
EV	EV	EV	EV
EW	EW	EW	EW
EX	EX	EX	

[illegible]

QY 361 TAXITAAVLTICHOBYLRTKTOAISKGRRLGVEHAQKFTTRLYSWLEKSGRDYIPGRQ 420
 Db 370 TAVITAAVLTICHOBYLRTKTOAISKGRRLGVEHAQKFTTRLYSWLEKSGRDYIPGRQ 429
 QY 421 FYAQRRLMSAGFHLDPVLYDESVPCRCRFLKLVAGKFCCEFRMWLGQECTCLLEPAE 480
 Db 430 FYAQRRLMSAGFHLDPVLYDESVPCRCRFLKLVAGKFCCEFRMWLGQECTCLLEPAE 489
 QY 481 GLVGHGHDNEAESESEVDPAPAHLDVGYAVYAGHOLEALYRLANTYPODIAAASRLT 540
 Db 490 GAVGQDGHNEAESESEVDPAPAHLDVGYAVYAGHOLEALYRLANTYPODIAAASRLT 549
 QY 541 ATVELVASPDLRECFRTVLGKNTFRTTVDGAHLLEANGPEQYVLYSPDARSQSGASHLT 600
 Db 550 ATVELVASPDLRECFRTVLGKNTFRTTVDGAHLLEANGPEQYVLYSPDARSQSGASHLT 609
 QY 601 YELTPAGIQRVSSNGLCTATFPFGCAPSAAPGEVAAPCSALYKYNFTQHSITGLGM 660
 Db 610 YELTPAGIQRVSSNGLCTATFPFGCAPSAAPGEVAAPCSALYKYNFTQHSITGLGM 669
 QY 661 LHPGELGIFPPEFSGHIMESANPFCGGLTYTRTWS-TSGFSSDFSP----PEAAPA 714
 Db 670 FHPEGLIGIFAPISPGHWEANPFCGGLTYTRTWS-TSGFSSDFSP----PEAAPA 729
 QY 715 MAATPGLPHSTPPVSDIWLPPPESEFQVDAVY-PPAPDPAGLP-GPVLTTPPPPPVH 772
 Db 720 MAATPGLPHSTPPVSDIWLPPPESEFQVDAVY-PPAPDPAGLP-GPVLTTPPPPPVH 776
 QY 773 KPSIP-PPSRNRRLYTPDCAKAVYAGSLFESDCMVLNANSPGHRPGGGLCHATYQRP 831
 Db 767 KPSIP-PPSRNRRLYTPDCAKAVYAGSLFESDCMVLNANSPGHRPGGGLCHATYQRP 826
 QY 832 EAFYTEEFIMREGLAAYLTLPRIITHAVALDYRVQNPRLAARFECRSRGTAAVPL 891
 Db 827 EAFYTEEFIMREGLAAYLTLPRIITHAVALDYRVQNPRLAARFECRSRGTAAVPL 886
 QY 892 GSGIYQVVSLSFDMERNHRPGDELTYLPEAANFEPANPQPLTTEDTARTANLAL 951
 Db 887 GSGIYQVVSLSFDMERNHRPGDELTYLPEAANFEPANPQPLTTEDTARTANLAL 946
 QY 952 ELDAATEVGRACAGCTISPGIYHOFYTAGVPSGSKRSIQGCDVYVYVPTRELNSMR 1011
 Db 947 ELDAATEVGRACAGCTISPGIYHOFYTAGVPSGSKRSIQGCDVYVYVPTRELNSMR 1006
 QY 1012 RGFAPFPHHTARVYIGRRVYIDEAPSLPHLLILHMRASSVHLGSPNQIPADFEHA 1071
 Db 1007 RGFAPFPHHTARVYIGRRVYIDEAPSLPHLLILHMRASSVHLGSPNQIPADFEHA 1066
 QY 1072 GLVPAIRBELAPTSMWXYTHRCPADVCELIRGAYPKIQTTSRVLSLFWNEPAIGOKLVX 1131
 Db 1067 GLVPAIRBELAPTSMWXYTHRCPADVCELIRGAYPKIQTTSRVLSLFWNEPAIGOKLVX 1126
 QY 1132 TOAKAANPAGATVHEAOGATFETTTATADARGLTOSSRAHAIVALTRHEKCVIIDA 1191
 Db 1127 TOAKAANPAGATVHEAOGATFETTTATADARGLTOSSRAHAIVALTRHEKCVIIDA 1186
 QY 1192 PGLIREVIGISDIYNNFLAGGEVXHRPSVTPRGPNQONIGTLOAFPPSOIASYHQ 1251
 Db 1187 PGLIREVIGISDIYNNFLAGGEVXHRPSVTPRGPNQONIGTLOAFPPSOIASYHQ 1246
 QY 1252 EELGHRPAPVAVLPPCELEGLYMPQELTVSDVLFELTIDVHMRMAAPSORKAVL 1311
 Db 1247 EELGHRPAPVAVLPPCELEGLYMPQELTVSDVLFELTIDVHMRMAAPSORKAVL 1306
 QY 1312 STLVGRYGRKRLYEAAHSDVRESIARFIPTIGVQATTCCLYELVEAMVEKQDGSV 1371
 Db 1307 STLVGRYGRKRLYEAAHSDVRESIARFIPTIGVQATTCCLYELVEAMVEKQDGSV 1366
 QY 1372 ELIDCNBVSRTFROKCKNFTTGEFTAHGKVGOGISAKSTFCALGPMFRAIEREL 1431
 Db 1367 ELIDCNBVSRTFROKCKNFTTGEFTAHGKVGOGISAKSTFCALGPMFRAIEREL 1426
 QY 1432 ALPPNIFYGAYEESVFAAVVSGAGSCMFENDESEFSDQNNNSTLGLCVMECCMP 1491

Db 1427 allpgvfygdafddtvsaaavaakamvfendfsefsgnfsjglecaineecgmp 1486
 QY 1492 QMLIRLYHLVRSAMTLOPKRSIKGFMKHSGEPTLIMVNMNATIAHCYERFDPVFA 1551
 Db 1487 QMLIRLYHLVRSAMTLOPKRSIKGFMKHSGEPTLIMVNMNATIAHCYERFDPVFA 1546
 QY 1552 AFKGDSDVYLCSDYKRSNMAALJAGCGLKLVDRPRLGYAGVAVPAGLTPDVPVFA 1611
 Db 1547 AFKGDSDVYLCSDYKRSNMAALJAGCGLKLVDRPRLGYAGVAVPAGLTPDVPVFA 1606
 QY 1612 GRLEKKNKGPBERAOLRLAVCDPLRGLTNYAOVCVDVSVRYGVSGLYHNLGMLOT 1671
 Db 1607 GRLEKKNKGPBERAOLRLAVCDPLRGLTNYAOVCVDVSVRYGVSGLYHNLGMLOT 1666
 QY 1672 IADCKAHFTETIKPVLDLTNSIORVE 1698
 Db 1667 IADCKAHFTETIKPVLDLTNSIORVE 1693
 RESULT 8
 AAW81519
 ID AAW81519 standard; Protein: 1693 AA.
 AC
 XX AAW81519;
 DE 02-FEB-1999 (first entry)
 DE Hepatitis E virus (HEV) ORF-1 protein.
 XX
 DE Hepatitis E virus; HEV, SAR-55; diagnostic agent; vaccine; antibody;
 KW passive immunisation; open reading frame; ORF.
 KW
 XX Hepatitis E virus.
 OS
 XX
 FH Key Location/Qualifiers
 FT MISC-difference 1238
 FT /note= "encoded by CAG"
 FT MISC-difference 1244
 FT /note= "encoded by CAG"
 FT MISC-difference 1352
 FT /note= "encoded by GTG"
 FT MISC-difference 1662
 FT /note= "encoded by GGC"
 XX
 EN W09846761-A1.
 XX
 PD 22-OCT-1998.
 PD
 XX 09-APR-1998; 98WO-US07418.
 PF
 XX 11-APR-1997; 97US-0840316.
 ER
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Emerson SU, Purcell RH, Robinson RA, Tsarev SA.
 PI WPL: 1998-568733/48.
 DR N-PSDB: AAV71604.
 XX
 XX New hepatitis E virus DNA from Pakistani strain SAR-55 - used for,
 PT e.g. developing products for diagnosis of, and vaccination against
 PT hepatitis E virus infection
 XX
 XX Disclosure; Pages 11-15; 204pp; English.
 PS
 XX
 CC This represents a hepatitis E virus (HEV) open reading frame (ORF)-1
 CC protein encoded by a DNA sequence designated SAR-55. SAR-55 also encodes
 CC HEV ORF-2 and ORF-3 proteins. A host organism transformed or transfected
 CC with a recombinant expression vector containing the SAR-55 nucleic acid
 CC can be used to produce the HEV proteins, especially ORF-2 protein. The
 CC recombinant HEV proteins can be used as diagnostic agents and as vaccines
 CC for use against HEV infection. The detection of antibodies specific for

CC HEV can be used for the diagnosis of infection and diseases caused by
 CC HEV, and for monitoring the progression of such disease. Such methods are
 CC also useful for monitoring the efficacy of therapeutic agents during the
 CC course of treatment of HEV infection and disease in a mammal. The
 CC antibodies can be used for detection or for passive immunisation of
 CC mammals.

XX Sequence 1693 AA;

Query Match 82.5%; Score 7439; DB 19; Length 1693;
 Best Local Similarity 81.6%; Pred. No. 0;
 Matches 1397; Conservative 111; Mismatches 162; Indels 42; Gaps 7;

QY 1 PGTTTIEQALAAASALANAVVPELSRQTELLINLMOPROLVFREVLMNHPTOR 60
 DB 10 pglitalegaalaanalanavvripfshqjellimqprqlvifrevlmnhpqr 69
 QY 61 VINNELEYCARAGKLEYGAPHSINDPNVLRHCFELRPVGRDVQRMYSAPTRGPAN 120
 DB 70 vlnneleycaragrcleigahprsinhpnvhrcflrpagrdvqrmysaptrgpaan 129
 QY 121 CRSSALRGLEPADRYTCDFGFSRCAPAEETVALYSLHDLMPADVAMARHGXTRLYAA 180
 DB 130 crssalrglpadrcyfdgfsrgcnpaetgialyslhdmspdvaeemftrhgmtrlyaa 189
 QY 181 LHPPEVLAPGTYHTTSYLLIHSDRAVVTYEGDTSAGCNHNVSTLRARIRTKTYGDH 240
 DB 190 lhppevlappgtyhttsyallindgtrvvtyegdtsagcnhvnstlrvstlrvtktygdh 249
 QY 241 PLVIEVERAIGCHFVLLTNAPEPSPMPYVPYRSTEVYRSIFGPGSGSLTFPSACSTK 300
 DB 250 plviertraigchfvllltaapepmpypyrstevyrslfpggtpsltfpscsk 309
 QY 301 STRHANPVHIMRLMFGATLDDQAFCCSRMLMTYLRGISTKVTGALVANEGMNASEDAL 360
 DB 310 sthvanpahlmrlmfgatlddqafccsrmlmtylrgistkvtvgalvanegmnasedal 369
 QY 361 TATTAAYLTICHORFLRQAIKGMRLGVEHAOKPTITLWLTPEKSGROVITPGRLO 420
 DB 370 taitaayltichqrlrqaiksgmrlerehaqfllrlyswltxesrdrlypqrql 429
 QY 421 FYACRRMLTSAGFHLDPRLVLPDESVPCCRFLKRYAKGFCCEFMRLAGECFCLEPAE 480
 DB 430 fyagcrrmltsagfhlprvlvdesapchortairkavskfcfcfmkvgceccflgpa 489
 QY 481 GLVGDGHNDNEATGESVDPAPBAHLDVSGTVAHGHOLEALYBALNVPODIARASRLT 540
 DB 490 glvgdghndneatgesvdpbaesaalsdsgsyvvgpialqplpaelpaivaragrlt 549
 QY 541 ANVELVASPDRLCETRVILGKTRFTYVDSAGHLEANGPBOYVLSFSDASRQSMAGSHSLT 600
 DB 550 atkvsvdgvdgrldcetrlgktrfsvdgavletnqperhnlstfdaagstmaagpfslt 609
 QY 601 YELTPAGLOVRSSNGLDCTATFPFGCAPSAAGEVAAFCFSAFYRNRFQSRSLTGLTW 660
 DB 610 yaasaglevryaaagldhnavfapgyvsrpagervtafsalyrfrnregrslgnfw 669
 QY 661 LHPBGLLGTFPPSPGHIMESANPFCGEGTLTYRTWS-----TSFGSSDSFSPPE 709
 DB 670 fhpegllgfpfapspghivesanpfcgegtlyrtlwesvdavpspdpqldgfs-----e 724
 QY 710 AAAPMAAATPGLRHSTPPPSDITWLRPPSEEFVDAAPV-PPAPDDA-GLPGVVLTPPP 767
 DB 725 pspisaaap-CP-----aaaplpppadpsptlsapargpap 761
 QY 768 PPEVHAKPSIP-PPSRNRRLITYPDGAKYVAGSLFESDCDWLVANANSPGRPGGLCHAF 826
 DB 762 gatarapaltthgatarhrlfitypdsksvtaglsfctcwlvaanavdhirpvgglchaf 821
 QY 827 YQFPEAFYTFEIMREGLAAYTLPRPIIHAAVPDYRVEONDKRLAAVRETCSRGTVA 886
 DB 822 yqfpeasfdaasfvmrdaaaylltprpiihavapdyrlenhpkrlaaeyretcsrlgta 881

QY 887 APYLLSGITQVAVVSLSFDAWERNHRPGDELXYLTERPAANFPAKRAQVLTITTEDTART 946
 DB 882 apyllsgityqvavvslsfdaawernhrpgdelylterpaanfpakraqvltitledart 941
 QY 947 ANLALEIDATEYGRACAGCTISPGIVHYQFTAGVPSPGSGSRSLIOGDVVVVVPTRELR 1006
 DB 942 anlaleidateygracagctispgivhyqftagvpspgsgsrsltqgdavvvvptrelr 1001
 QY 1007 NSWRRGFAAFPTHTAARVTIGRRVVIDEAPSRLPHLLLMORASSVHLLGDPNQIPAI 1066
 DB 1002 nswrrgfiafpthtaarvtrgrvvvdeapsrlphlllmgraatvhlhgdpnqipai 1061
 QY 1067 DEFHAGIVPAIRPELAFTSMWXYTHRCAPAVCEILINGAYKICTTSVYLSLFWNNEPATG 1126
 DB 1062 dfhagivpaiprelaftsmwxythrcapavcelingaykicttsvylslfwnepatg 1121
 QY 1127 QKLVXTQAAKAPNGAIVTHEAGATFETETIATADARGLIOSSRAHAIVALTTRHEKC 1186
 DB 1122 qklvftqaaapngavtvehaagalyteltliatadargllqsrthavalttrhkekc 1181
 QY 1187 VILDAPGLREVGISDVYVNNFFLAGGVGXHRRSVIIPRGNDQNTGTLQAPPPSQISA 1246
 DB 1182 vildapglirevgisdaivnnfflaggvgxhrrsviiprgnpandtlaafppscisa 1241
 QY 1247 YHGLAEELGHRAPAVAVLPCELEOGILMYRQELTVSSVYVPELTDVHCRMARPSQ 1306
 DB 1242 yhglaelghrapavaavlpcelegillyrpeeltcdsvfelftdlvhormarpsq 1301
 QY 1307 RKAVALSTLVGRYGRRTKLYFAAHSADVESLARTIPRTGPQATTCETLYELVEAMVEKGD 1366
 DB 1302 rkavalstlvgryrtrklyfaashsdvrsdarflpaaipgvqvtcelyleeamvekgd 1361
 QY 1367 GSAVLELDLCNRDVSRTTFQKXCNKFTTGETIAHGKVGOGISAMSKTCALFGPMFRAT 1426
 DB 1362 gsavleldlcnsrdvsrttfqkcnkfttgeltahgvvggisaawsktfcalfgpfrat 1421
 QY 1427 EKELIALPPNIFRYGAVYEEVSFAAASVSGSGMVENDESEFDSQNNNSLGLCEVYME 1486
 DB 1422 ekelialppnifrygavyyeefsaasvsgsgmvendesefdsqnnnsldcevy 1481
 QY 1487 ECGMPQMLIRLYHIVASAMTILQAPKESLKGFWKHSSEPTLIMTWNNAAIIAHGCEFR 1546
 DB 1482 ecgmpqmlirlyhlisawllqapkeslrgfwkhsgeptllmwtvwnmaavltthyodfr 1541
 QY 1547 DFRVAAFKGDSVYVLCSDYRQSRNAALLAGCGLKLVDRYRPIGLYAGVVAPGLSTLPD 1606
 DB 1542 dfrvaafkgdsivylcseyrqpagaavllagcgllklkvdrpigliyagvvapglstlpd 1601
 QY 1607 VVRFAGRLSEKMGPGPERAEORLAVCDRLRGLTENVAVQVDSRVYVSGVGLVHNLT 1666
 DB 1602 vvrfagrlsekmgpgperaeqrlravsdftrkltnvaqmcvdsrvyvgvsglvhnll 1661
 QY 1667 GMLQTTADGKAMHETETIKPVLDTNSIIORVE 1698
 DB 1662 emlqavdgkahlteevkpvldltlnslcive 1693

RESULT 9
 AAM76368
 ID AAM76368 standard; Protein; 1693 AA.
 XX
 AC AAM76368;
 DT
 XX
 XX 03-DEC-1998 (first entry)
 DE Hepatitis E virus hollow particle protein #1.
 KW hollow particle protein; virus; antibody; detection; immunoassay;
 XX infection.
 XX
 OS Hepatitis virus.
 XX

FH Key Location/Qualifiers
 FT Protein 1..1693
 FT /note="Partial sequence"
 XX JP10234383-A.
 XX
 XX 08-SEP-1998.
 PD
 XX
 XX 28-FEB-1997; 97JP-0062445.
 PF
 XX 28-FEB-1997; 97JP-0062445.
 PR
 XX (DENK-) DENKA SEIKEN KK.
 PA (KOKU-) KOKURITSU YOBO EISEI KENKYUSHO.
 XX
 DR WPI: 1998-535037/46.
 XX N-PSDB: AAV61687.
 XX
 PT Hepatitis E virus hollow particle poly:peptide(s) and nucleic acids
 PT encoding it - useful for more accurate detection of HEV in samples,
 PT using immuno-assays and nucleic acid hybridisation
 PS
 PS Claim 10; Page 17-24; 29pp; Japanese.
 CC This sequence represents a Hepatitis E viral hollow particle protein.
 CC This polypeptides can be used to raise antibodies to detect HEV
 CC infection in samples, e.g. by immuno-assay based techniques, and the
 CC nucleic acid can be used for the same in nucleic acid hybridisation
 CC assays. The polypeptides and nucleic acids allow more accurate
 CC detection of HEV than previously possible.
 CC
 XX
 SQ Sequence 1693 AA;

Query Match 82.5%; Score 7437; DB 19; Length 1693;
 Best Local Similarity 80.9%; Pred. No. 0;
 Matches 1391; Conservative 114; Mismatches 158; Indels 56; Gaps 5;

QY 1 PGITTAIEGAAALAAANSALANAVVPELRSVOTELINLMQPROLVEPEVLMNHPQR 60
 DB 10 piltalaeqaaalaanaaanaavvrflshqjellnlmmpqrqlvtrpevfwmhpr 69
 QY 61 VHNLEEQCRARAGKCEVEGAHPSRINDPNVLRHRCFLRPVGRDQWMSAPTRGPAAN 120
 DB 70 vhnleelycraragrcleigahpsrindpnvnrhrcflrprgrdqrtyrtepttgrpaan 129
 QY 121 GRRSALRGLPADRYTCDFGSRCAFAETGVALYSLHDLMPADVAEAMARHGXTRLYAA 180
 DB 130 grrsalrglpadvdrtycldgfsgrcfpaetgialyslhdmsspsdvaeamfrngmtrlyaa 189
 QY 181 LHLPEVLLPQGTHTTSTVLLIHDGRNAVTEGDRSAGYNHDVSLRAMITRTKIVGDH 240
 DB 190 lhlpevllppgcytlaasylllhdgrnavvtegdtsagynhdvsnllswitrltkvgdh 249
 QY 241 PLVIERVRAIGCHFVLLTLAAPESPMPVPPYRSTEVVRSIFGPGSSPLFSPASCSTK 300
 DB 250 plviervraighcfvllllaapespmppyprstevvrsifpggspslfpscscstx 309
 QY 301 SFTHAVPNIWIRLMFEGATLDDQAFCCSRMLTYLRGISYKTVGALVANEGMNASDIAL 360
 DB 310 sfthavpahniwdrllmfegatlddgaafccsrmltylrgisylkvtvgtlvanegmasedial 369
 QY 361 TFXITAAVLTICHOIRLFRQATISGKMRIRVGEHAKFIRLRLSMLEKSGROYITGRORO 420
 DB 370 tavitaaayltichgrlylfrqatlskgmrirvgehaqffirplyswlfeksgroyitgr 429
 QY 421 FYAQCRRMLASGFHLDPRLVLFDESVPCRRTEFLKVKAGFCCEFMKMLGOECTCELEPAE 480
 DB 430 fyaqcrmlasgfhldprlvlfdesvpcrrteflkalksfccfmkwlgectcflgpaee 489
 QY 481 GLVGHGHDNEAYEGSEVDPAEPAHLDVSGTYAVHGHOLEALYRALNVPQDITAAARSRLT 540
 DB 490 glvghghdneayegsevdpaepahldvsgtyavhgholealylralnvpqditaaarsrlt 549

QY 541 ATVELVASEDRLECRTEVLGNKTRFTTVYDGAHLAENGPEQYVLSFDASQSGASHSLT 600
 DB 550 atvkvsgdgridceltllgnktrftrtsfvdgavleangperylslsdasqstmaagpslpt 609
 QY 601 YELTPAGLOVRISSNGLDCTATFPPGAPSAAPGEVAAFCSLYRXNRTQHSITGGIM 660
 DB 610 yaasaaglevryvaagldhravfapysprsaagevtalcslalryfnreaqhsitgimlv 669
 QY 661 LHPGGLGIFPPSPGHIMESANPCGECTLYTRTWS-TSGSSDPSF-----PAAAPA 714
 DB 670 fhpegllglfapfsgphwesanpcogectlytrtwsevdavsparpdlgmsepslps 729
 QY 715 MAATPGLPHSPVSDIWLPPPSSEFOYDAAPVPAPDAGLPGPVYLTPPPPVHKKP 774
 DB 730 raatpctl-----aaplp-----lapdpppppsap 754
 QY 775 STPPP-----SRNRRLTYTPDGAKVYAGSLFESDCDMLVNASNGHRPG 819
 DB 755 aldepasgatagapaltqtharhrllyfcpdgskvfagslfesactwlvnasndhrpg 814
 QY 820 GGLCAFYQRPPEAAYPFEMIRREGIAAVTLPRPIIHAVADYVEQMPKRLAEAYREPT 879
 DB 815 gglcayfyqrpasidaasfvmrdgaayclcpriihavadyrlenmpkrlaeaayrec 874
 QY 880 CSRRTAAVPLIGSGIYQVPVLSFSDAMERNHRPDELYTPPAAWFEANKPAPVLTIT 939
 DB 875 csrltaayplligtlyqvpvlsfswdamerhrpdeplytpalaarwfearnpctrltli 934
 QY 940 TEDTARTANLALEIDAATEVGRACAGCTISPGIVHVOFTAGVPSGSKRSIQGDVDVYV 999
 DB 935 tedvartanlaleidaatevgracagctispgivhvoftagvpsgskrsitqadvdvvy 994
 QY 1000 VPTRELRNSMRGRFAAFTPHAAAVTIGRRVVIDEAPSLPRHLLLMHQRASSVHLIGD 1059
 DB 995 vptrelrnsmrrgrfaaftphtaavtigrvrvideapslphllllmqraavhlligd 1054
 QY 1060 PQQIPAIIDFEHAGLVPALIRELAPTSWMXVTHRCPADVCCELRCGAVPKIOTSRVLSIF 1119
 DB 1055 pqqipaidfehaglvpalirpdlaptswmvthrcpadvccllrgaypmqttsrvlsif 1114
 QY 1120 WNEPAIGOLVYTOAKAKANPGATVHEAGATFETTTIATADARGLIOSSRAHAIVAL 1179
 DB 1115 wgepavgqlvftgaakaangpsvtvheagatfettliatadargliqssrahaival 1174
 QY 1180 TRHTEKCVILDPAGLLREVGISDVYVNNFTLAGGEVGHAKRPSVIRGNNDQWLGLQAF 1239
 DB 1175 trhteksvildasgllrevgsdaivnmfllaagelgqirpsvlpngpndavndclaafr 1234
 QY 1240 PSCQISAVYHQLAEELGHRPAPVAAVLPCCPELEOGLIYVPOELATVSDSVLVELELDIYHC 1299
 DB 1235 pscqisafhqlaeelghrpavvaavlpccpelvqgllylpgqltaodsvvleldvnhc 1294
 QY 1300 RMAAPSORKAVLSTLVGRYGRRTKLYEAAHSDVRSLARFIPTIGPVQATCELYELVEA 1359
 DB 1295 rmaapngktavlstlvgryrtrtklynashsdvrsldarflpajpgvqvttcellyelvea 1354
 QY 1360 WYKGGODGSAVLELDLONDNVSRITFPQKXCKFTTGFIANGKVGIGSASTFCALF 1419
 DB 1355 mwkkgodgsavleldcnvdsvritfpgkdxckfttgeliaqkvvgysaastfcalf 1414
 QY 1420 GPMFRAIEKEIILALPNNIFGYDAYEESYFAAASGAGSCWFENDSFEDSTONNFSIG 1479
 DB 1415 gpmfraiekaiallppgvfygdadldcvsaavaakasmvfendfsefstqnfslg 1474
 QY 1480 IECVVMEECGMPQWLIRLYHLVRSAMLIQAPKESLKGFWKKHSGEPGLTLMNTVMNMAIT 1539
 DB 1475 lecvvmeeqmpqwlirlyhlrsawliqapkeslrgfwkksngsgpghtlmlntvmnmai 1534
 QY 1540 AHCFEPRDFRVAAFKGGDSVVLCSDYROSRNAALILACGGLKLYKDYPRIGIYAIVVAVP 1599
 DB 1535 thcyftrdlqvaaifgddslvscseyrqspsgaavllaagcklkvkdfrpdlgyavvvavp 1594

Query	Match	Score	DB	Length
Best Local Similarity	82.5%;	7437;	DB 19;	Length 1693;
	81.7%;	Pred. No. 0;		

```

Oy 1072 GLVPAIRRELATSMWXYTHRCPADVCELIRGAYPKIOTTSVLRSLFPMNEPAIGOKLVX 1131
Db 1067 GLVPAIRPDLPTSWHTRHPADVCELIRGAYPMIGTTSVLRSLFGEPAVQKLVF 1126
Oy 1132 TOAANKANPAGATVHEAGATFETTTIATADARGLIOSSRAHAVALTRHTEKCVILDA 1191
Db 1127 tgaakpangpsvthneagacytettitlatadarglqssrahaivaltrhlekcvilda 1186
Oy 1102 PGLLEVEGISDYVNNFPLAGEVGHKRPVYPRGNPQONLGLAOFPPSCQISAYHQLA 1251
Db 1187 PGLLEVEGISDAIVNFIAGAGELGHQPSVPIRGNPDANDVCLAAFPSCQISAFHQLA 1246
Oy 1252 BELGRHPAPVAALPPECLEGLLYMPOELTSDSVVFEFTDVIHORMAPSORKXAVL 1311
Db 1247 eelgrhpyvvaavlppecleeglllylpqeltdcsavtfeftldvthcmmapsqrkav 1306
Oy 1312 STLVRGRRKRLYEAAHSDVRESLAREPTIGVQATTCLELVEAMVERKQDGSAYL 1371
Db 1307 stlvgrgygrkllynashsdvreslarfipaigpyvttccllyelveamvekgdgsayl 1366
Oy 1372 ELDDCNROVSRITFFQKCNKRTTGETTAHGVGOGISMSKTFPCLGPMFRALKEKEL 1431
Db 1367 elddcnrvsriftffqkcnkrttgettlahgvvggisawsktfcallgpmfralekell 1426
Oy 1432 ALPPIFYGDAYEESVFAAASGAGSCVFNDESEFSTQNNFSLGLECVAMEECGMP 1491
Db 1427 alppigvfygdafddtvsaaavaakamvfndesefstqnnfslglecvameecgmp 1486
Oy 1492 QWLIRLYHLVRSAMILLQAPKESLKGFWKKHSGEPCTLLMNTVNMNAIIAHCEYERPRVA 1551
Db 1487 qwlirlyhlvrsawillqapkeslrgfwkxhsgepctllmvtwnmaavilthcydfdfgya 1546
Oy 1552 AFKGDSDVLCSDYOSNRMAALINGCGIKLVDRPRTGLVGVVAVAGTGLPVPVRA 1611
Db 1547 afkgxsavlcsytrgspgaavliagcgiklvdrpirlgylvagvavapglpavvra 1606
Oy 1612 GLTSLKNNMGCPGERAEQRLAIVCDPLRGLTNYAQVCVDVSVHVSFGLVHNLIGMLQT 1671
Db 1607 gltcknngpgeraeqrlavsdflrkltnvaqmcvdvsvhyvsgplvnhnligmqla 1666
Oy 1672 IADGKAHETETKPYVLDLTNSTIIOQVE 1698
Db 1667 vadgkahfetsvkvyltltnsilcrve 1693

RESULT 11
ID AAR51264 standard; Protein: 1693 AA.
AC AAR51264;
XX 21-OCT-1994 (first entry)
Dt
XX HEV strain protein encoded by ORF-1.
De
XX Hepatitis E virus; HEV; strain SAR-55; open reading frame; ORF;
KW antibody; detection; diagnosis; primates; stool suspension.
XX
XX Hepatitis E virus strain SAR-55.
OS
XX WO9406913-A.
XX
XX 31-MAR-1994.
PD
XX 17-SEP-1993; 93WO-US08849.
PF
XX 18-SEP-1992; 92US-0947263.
PR
XX (USSH ) US SEC DEPT HEALTH.
PA
XX Emerson SU, Purcell RH, Tsarev SA;
XX

```

```

DR WPI: 1994-118462/14.
DR N-PSDB: AA045197.
XX
XX Purified hepatitis E strain SAR-55 virus - used to develop prods.
PT for use in detection, diagnosis, vaccines and therapy of
PT hepatitis E virus infection
PS
XX Disclosure; Page 57-62; 114pp; English.
XX
XX The sequences given in AAR51264-66 are encoded by the hepatitis E virus
CC (HEV) strain SAR-55. The cDNA sequence contains three open reading
CC frames (ORFs). These proteins can be used to stimulate the production
CC of protective antibodies upon injection into a mammal that would serve
CC to protect the mammal upon challenge with wild type HEV. The proteins
CC can be used for detection and diagnosis of HEV infection. The HEV
CC SAR-55 cDNA was isolated from primates inoculated with stool
CC suspensions obtained from hepatitis E patients.
XX
XX Sequence 1693 AA:
SQ

```

```

Query Match 82.5%; Score 7436; DB 15; Length 1693;
Best Local Similarity 81.5%; Pred. No. 0; Mismatches 162; Indels 42; Gaps 7;
Matches 1396; Conservative 112;

```

```

Oy 1 PGITTAIBQAAIAANSAANAVVVRPFLSRVOTETILINLMOPROLVPRPEVLMNHPJOR 60
Db 10 pgitaiqbqaaiaaansalanavvrflsrvgqellinlmopqrlvtrpevfwhpqr 69
Oy 61 VIHNELEOYCARAGRCLEVGCAHPRSINDPNVHLRCLRPYGRGVOWYSAPTRGPAN 120
Db 70 vihneleoycarsgrclevgahprshndpnvhlrcrlrpygrgvowysapttrgpan 129
Oy 121 CRRSALRGLPADRYCYDGRFCRCAFAAETGALXSLHDLMPADVAEAMARHGXTRYAA 180
Db 130 crrsalrglpadrycydgrfrcrcafaaetgalslhdmpadvaaeamrfgxtryaa 189
Oy 181 LHLPEVLLPRTYHTTYSYLLIHDDRAVVTYEGDTSAGVNHDSILRAMIRTTKIYGDH 240
Db 190 lhlpevllprrtyhttyllyihdgrvvtlyegdtsgaynhdvanlswirtkltygdh 249
Oy 241 PLYIERVRAICGFVLLTAAPESPMPYVPRSTEVYVNSIFRPGSPSLFPACSTK 300
Db 250 plyiervraicghfvlltlaapepmpyvprrstevyvnslfrpgspslfprscstk 309
Oy 301 STEFHAVPIHMDRLMFCGATLDDOAFCCSRIMTYLRGISYKYTVGALVANEGMNASDAL 360
Db 310 stefhavpihmdrlmfcgatlddgaafccsrilmtylrghsykytvglvanegmasedal 369
Oy 361 TAXITAAVLTICHOHYLRTQALISKGMRLGVENHAKFTTRLYSWLFEKSGRDYIPGROLQ 420
Db 370 taxitaaavlthohorylrtqaliskgmrlgvenhakfttrlyswlfeksgrdyipgrgle 429
Oy 421 FYAGCRRLWSAGFLDPRVLYFDESVPCKRFLKLVKAGKFCFCFMRMTGQETCGLPFAE 480
Db 430 fyagcrrwlsagflhprvlyfdesaapchrtalkavskfcfcmrkwtgqetctclpfae 489
Oy 481 GLVGDHGNDAVEGSEVDPAEPALHDVSGTYAVGHOLEALRYALNPOTIARASRLT 540
Db 490 glvgdghndavegsevdpaepalhdvsgtyavgholealryalnpotiaraasrlt 549
Oy 541 ATVELVASPDLRCKRTVLGNKTFRTTVVDGAHLNANGPEQYVLSFASROSQKAGSHSLT 600
Db 550 atvkvsgdygridcettlgnktrftrtvvdgahlnangpelyvlsfdaasqstmaagrsalt 609
Oy 601 YELTPAGLQVRISSNGLDCTATPPGAPASAPAGVAAFCMALRYKNFTQHSITGGLM 660
Db 610 yaasaaglevryvaaglhtrvafapgvprsapgvtafosalylfnreaqslstgnfw 669
Oy 661 LHPGLIGIFPSPSGHIMESANPRCGGLTYTRFWS-----TSGSSPFSPE 709
Db 670 fhpegllgpfapfsgphwesanprcgstlytrtwsevdavspapqdlgfts-----e 724

```

```

QY 710 AAAAAAAAAAATPGJPHSTPVSIDWLPPESEFQVDAAPV-PPAPDBA-GLPGFVLTTPP 767
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 725 psipsnaatp-----tp-----aapipppdpdpstplsaparegap 761

QY 768 PPRVHKPSID-PPSRNRRLIYTPDGAKVYAGSLFESDCDWLVNASNGHRPGGLCHAF 826
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 762 gataapaltghcartrrllftypdgskvfagslfestcwlvaasnvdhrpgglchaf 821

QY 827 YQFPPAFYPTFEIMREGIAAATLTPRIITHAAPDYRVQNKRLLEAARECSRGTA 886
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 822 yqfypasfdaastvmrmdgaaycllcpriihavapdyrlenhpkrlleaayretcslrgta 881

QY 887 AYPPLSGIYOYVSLSPFAMERNHRPGDELYTEPAAMFEANKPAQPLTITTEPART 946
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 882 ayplltgytqyvpigpsfdaewerhnrgdelypelaarwfeanrptcpllttedart 941

QY 947 ANIALEIDATEYGRACAGCTISPGIVHQTAGVPGSGKSRSIQGGDDVVVVPTRELR 1006
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 942 anialeidatdygracagcrtvpgvvyqftagvpgsgsrsltgadvvvvptrelr 1001

QY 1007 NSMRRGFAFPPTHTAARTIGRRVYIDEAPSLPHLLLLHMQRASSVHLGDPNOIPAT 1066
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1002 nawrtqfifaatfcphtaartvqgrvvideaipphllllhmqrattvhlhgdpnqpal 1061

QY 1067 DEEHAGLVPAIRPELAPTSMWXYTHRCPADVCELIGAYPKIQTTSRVLSLFWMNEPATG 1126
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1062 dfenagltypairpdiaptswmhwthrcpadvcclirgajpmigttervirslrtwgepay 1121

QY 1127 QKLVYQAQKAAANPGAIVVHEAGATFTETTLIATADARGLIQSSRAHAIVALTRTEKC 1186
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1122 qkvlftqaakaangsvttheagagalytettliatadargllqgsrahaivaltrtkec 1181

QY 1187 VILIDAGGLRENGISDVIYANNFPLAGGEVGHXHPSVITPKNPNQNGTLQAFPPSCQISA 1246
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1182 vildagglrtreygisdaivnnfflaggeglhngpvsviprgnpdanadtlaafppscseisa 1241

QY 1247 YHOLABELGHRPAVAAPVPCPELEOGLLYMPQELVTSVYVFELTDIVHORMAPSQ 1306
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1242 fhelaeelghrpepaavvlppecelegallylpqeltdcsvvtfeldvthcrmaapsg 1301

QY 1307 RRAVISTLVGRYGRRTKLYEAHSDVRESLARPIPTIGVQATTCELYELVEAMVEKGD 1366
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1302 rkavistlvgryrtrklynaashdvrdslarlipalgrqvvtcelayelaeamvekqgd 1361

QY 1367 GSAVLELDLCNRDVSRTTFQKKXKFTTGTTAHGKVGGGISAMSKTFCALFGWPRAT 1426
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1362 gsavleldlcsrtvsrtlftqkcnkfttgetlahgkyvggisaawsktfcalfgwpfrral 1421

QY 1427 EKEITALLPPIFYGDAYESVFAAIVSGSGMVFENFSEPDSDONNPSLGLCEVME 1486
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1422 ekaitallpbgvtygdaiddvtsaavaakaasmvtendfsetscqnhtslglectime 1481

QY 1487 ECGMPWLRLYLIVRSANILQAPKESLKGFMKHSGEPTLLMNTVMNNAIIAHCEYER 1546
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1482 ecgmpwlrllyllirsavllgqkeslrgfwkksgepglltwlvtvmnavllthoydfr 1541

QY 1547 DRRVAAFKGDSDVYLCSDYRQSRNAALALAGCGLKLKVDRIPIGLYAGVVAPAGLTLD 1606
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1542 drrvaafkqddslvlycseyrpsgaavllagcgllkvdfrpigliyagvvapaglgald 1601

QY 1607 VYRFAPARLSEKMGPGREABEOLRLAVCDFELRGTLTNVAOVYVRYVSGVSLVHNL 1666
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1602 vyrfaparlsekmgpgreabegrlavsdflrlkltmvaqmcvavsrlyvgsbplvhnll 1661

QY 1667 GMLQTIADKAHFTETIKRPVLDLTNSIIORVE 1698
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1662 emlgvadgkahnftesvkpvlldltmslcrle 1693

```

RESULT 12
 AAR14618
 ID AAR14618 standard; Protein; 1693 AA.
 XX

```

AC AAR14618;
XX
DT 16-JAN-1992 (first entry)
XX
DE Protein encoded by ORF 1 of Burmese ET-NANB viral strain.
XX
DE enterically transmitted non-A, non-B hepatitis virus; hepatitis C;
XX HCV; E.coli strain B84; ATCC deposit number 67717; Burma.
XX
OS Enterically transmitted non-A, non-B hepatitis virus.
XX
PN WO9115603-A.
XX
PD 17-OCT-1991.
XX
PE 05-APR-1991; 91MO-US02368.
XX
PR 05-APR-1990; 90US-0505888.
XX
PA (GENE-) GENELABS INC.
XX (USSR ) US DEPT HEALTH & HUMAN.
XX
PI Reyes GR, Yarbough PO, Bradley DW, Krawczynski KZ, Tam A;
PI Fry KE;
XX
DR WPI: 1991-325242/44.
XX N-PSDB; AAO14412.
XX
XX
XX New viral proteins from non A-non-B hepatitis agent - used to
XX treat and prevent enterically-transmitted non-A non-B hepatitis
XX virus
XX
XX Disclosure: Page 15; 117pp; English.
XX
XX
XX A positive clone ETL.1 was identified in a library prepared from
XX bile of cynomolgus monkeys infected with the Burma strain of ET-NANB.
XX Both strands of ETL.1 were sequenced. Identity of the sequence with
XX sequences in etiologic agents has been confirmed by locating a
XX similar sequence in a viral strain isolated in Burma. This protein
XX is encoded by the longest ORF (ORF 1) of the Burma strain.
XX (See AAO14410 for ETL.1).
XX
XX
XX Sequence 1693 AA.
XX
XX
XX Query Match 82.1%; Score 7400; DB 12; Length 1693;
XX Best Local Similarity 81.1%; Pred. No. 0;
XX Matches 1388; Conservative 110; Mismatches 172; Indels 42; Gaps 6;
XX
XX 1 PGTTTAEQAAALAAANSALANAVVRPFLSRVOTETILINMQPROLVFRPEVLMNHPQR 60
XX : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 10 pgtitaqdaalaanaalanavvrpflshgqieillimqprqlvfrpevfmnbpqr 69

QY 61 VHNLEIYOCRAARGCLEVGARPSINDPNVTLHRCFLRPVGRDVOORWSAPTGRPAAN 120
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 70 vhnleilyocrargclevgarpsindpnvthrcflrpvgrdvovrwsaptrgpaan 129

QY 121 CRRSALRGIPPADRYTCFDFGSRCAFAETGVALYSLHDLMPADVAEAMARHGXTRLYAA 180
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 130 crrsalrglpaadrtyclgdfsgcnfpaetgialylshdmspdvegmfrgmtrryaa 189

QY 181 LHLPEVLLRPGTYHTTSYLLHDSGRAVYTEGDTSGAGYNHDSILRAMIRTKTVGDH 240
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 190 lhlpevllrpgtyrtyasyllhdsgrvvytyegdtsagynhdsilrswlrtrtkvtdh 249

QY 241 PLVTEVRVIRIGCFVLLTAAPESPMPYVVPYRSEVYVRSIFEGGSGSPSPFSACTRK 300
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 250 plvtevririgcfvlltaapepmpyvpypyrsevyvrsifeggspspfsactrk 309

QY 301 STFHAAVPIHWRMLFGATLDDQAFCCSRMLMTYLGISYKYVGAALVANEGMNASDAL 360
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 310 stfhavpahwtrmlfgatlddgaftccsrmlmtylglsykyvgaalvanegmnasadal 369

```

```

QY 361 TATITAAVLTICHOXYLFTQAIKSGMRRLGVBNHAKFTITRLYSWLFKSGRDYIFGRQLO 420
Db 370 TAVITAAVLTICHOXYLFTQAIKSGMRRLGVBNHAKFTITRLYSWLFKSGRDYIFGRQLO 429
QY 421 FYAOCRRRLSGFHLDPVILVDESVPCRCRFLKKVAGKCCCFMRWIGOEOTCLLEPAE 480
Db 430 FYAGCRRLSAGFHLDPVILVDESVPCRCRFLKKVAGKCCCFMRWIGOEOTCLLEPAE 489
QY 481 GLVGDHGNENAESESEVDEPAEPAHLDVSGTYAVHGHOLEALRYALANYPDIAARASLT 540
Db 490 GAVGQGHNEAESESEVDEPAEPAHLDVSGTYAVHGHOLEALRYALANYPDIAARASLT 549
QY 541 ATVELVASPDLECTVYGNKTFKTTVVDAHLEANGPEQVYLSDSARQSGASGASHLT 600
Db 550 ATVKVSGVGRICETLLIGNKTFKTTVVDAHLEANGPEQVYLSDSARQSGASGASHLT 609
QY 601 YELPAGLOVRISSNGLCCTATFPFGGAPSAAPGVAFCALYNYNFTORHSHLTGGM 660
Db 610 YAAASAAGLEVRYAAGLHRAVFAFGFSPRSPGDTALCSALYRINAEQTHSLIGNLW 669
QY 661 LHPEGLIGFPPFSPGHIMESANPPCGEGLTYTRTWS-TSGFSSDFSP----PEAAPA 714
Db 670 FHPEGILIGLFAFPSPGHIMESANPPCGEGLTYTRTWS-TSGFSSDFSP----PEAAPA 729
QY 715 MAATGCLPHSTPPVSDIWLPPPESEFOYDAAPV-PPAPDAGLPGVYLTP----- 766
Db 730 RAATPCL-----AAPLPPPADPSPSPAPSAAPSAAGATCG 766
QY 767 PPPVHKSPSPSPNRRLTYTPDGAVYAGSLFESDCCDMVNSNGHRPGGGLCHAF 826
Db 767 APAATHQ-----TATNRLLFTYDPSKVFAGSLFESDCCDMVNSNGHRPGGGLCHAF 821
QY 827 YORPEAFYPTFFIMREGILAAATLPRPIIHAAVDYREONPKRLEAAYRETCRRGTA 886
Db 832 YGYPASFDAASFVMDGAAYTLCPRIHAVAPDYLEHNPXKLEAAYRETCRRGTA 881
QY 887 AYPILGSGIYOVVSLSDAMERNHRPGDELYLTPAANWFANPAPQVLTITDTART 946
Db 882 AYPILGSGIYOVVSLSDAMERNHRPGDELYLTPAANWFANPAPQVLTITDTART 941
QY 947 ANLALEIDAAVTEGRACGCTISPGIYHVOFTAGVPGSGSKRSIOGGDVVVVPTREL 1006
Db 942 ANLALEIDAAVTEGRACGCTISPGIYHVOFTAGVPGSGSKRSIOGGDVVVVPTREL 1001
QY 1007 NSMRRRGFAPTPHTAARTYIGRRVVIDEAPSLPHLLHLHMQRASVHLLGDPNQIPAI 1066
Db 1002 NAWTRIGFAATFHTAARTYIGRRVVIDEAPSLPHLLHLHMQRASVHLLGDPNQIPAI 1061
QY 1067 DFEHAGLVPAIRPELAPTSWMAVYTHRCPADVCELIRGAVPKIQTTSRYLRSLFMNEPAIG 1126
Db 1062 DFEHAGLVPAIRPELAPTSWMAVYTHRCPADVCELIRGAVPKIQTTSRYLRSLFMNEPAIG 1121
QY 1127 OKLYTOAKKANPGAIVHVEOGATFETTTIATADARGLIQRSAHAYVLTHTKRC 1186
Db 1122 OKLYTOAKKANPGAIVHVEOGATFETTTIATADARGLIQRSAHAYVLTHTKRC 1181
QY 1187 VILDPAGLREYGISDVIVNNEFLAGEVGHKRPVIRGNPDQMLGTLOAFPPSCQISA 1246
Db 1182 VILDPAGLREYGISDVIVNNEFLAGEVGHKRPVIRGNPDQMLGTLOAFPPSCQISA 1241
QY 1247 YHQLAEELGHRPAPVAAVLPCELEGGILYPOELTVSDSVLVELFDYVHCMAAPSO 1306
Db 1242 YHQLAEELGHRPAPVAAVLPCELEGGILYPOELTVSDSVLVELFDYVHCMAAPSO 1301
QY 1307 KRAVSTIVGRGRRTKLYEAHSDVRESIAFTITIGPVQATTOCELEIVAMERKOD 1366
Db 1302 KRAVSTIVGRGRRTKLYEAHSDVRESIAFTITIGPVQATTOCELEIVAMERKOD 1361
QY 1367 GSAVELDLGNDRVIRTFEFOKXCNKFTTGERTIAGKVGOGISAKTFCALFGWFPAL 1426
Db 1362 GSAVELDLGNDRVIRTFEFOKXCNKFTTGERTIAGKVGOGISAKTFCALFGWFPAL 1421
QY 1427 EKELLALPPNIFYGDAAVEEYFAAASGAGSCWTFENDESEFSDTONNFSIGLECYWME 1486

```

```

Db 1422 EKALLIPGVIFYGDAFDVDFSAVAAKSMWFENDISEFDSIGNISIGLECAM 1481
QY 1487 ECGMPQWILIRLYHIVRSAMILOAPKESIKGFWMKSGCEPGLTMMWTVMNAIHCYEER 1546
Db 1482 ECGMPQWILIRLYHIVRSAMILOAPKESIKGFWMKSGCEPGLTMMWTVMNAIHCYEER 1541
QY 1547 DFRVAAEFGKDSVYLCSQYKOSRNAAALIACGGLKLYDYPILGYAGVVAAPGLTLPD 1606
Db 1542 DFRVAAEFGKDSVYLCSQYKOSRNAAALIACGGLKLYDYPILGYAGVVAAPGLTLPD 1601
QY 1607 VYRFAGRLEKNGKGPGRAROLRAVCDPLRGLTNVQVCVDVSRVYGVSPGLVHNL 1666
Db 1602 VYRFAGRLEKNGKGPGRAROLRAVCDPLRGLTNVQVCVDVSRVYGVSPGLVHNL 1661
QY 1667 GMLQTIADGKAHFTETIKPVLDLTHNSIIQRYE 1698
Db 1662 GMLQTIADGKAHFTETIKPVLDLTHNSIIQRYE 1693

```

RESULT 13

AA14615
ID AA14615 standard; Protein: 431 AA.

AA14615;

16-JAN-1992 (first entry)

Encoded by ORF 1 of ET-NANB clone ETL.1 "forward" strand.

enterically transmitted non-A, non-B hepatitis virus; hepatitis C; HCV; E.coli strain B84; ATCC deposit number 67717; liver disease.

Enterically transmitted non-A, non-B hepatitis virus.

W09115603-A.

17-OCT-1991.

05-APR-1991; 91WO-US02368.

05-APR-1990; 90US-0505888.

(GENE-) GENELABS INC.

(USSH) US DEPT HEALTH & HUMAN.

Reyes GR, Yarborough PO, Bradley DW, Krawczynski KZ, Tam A;

WPI; 1991-325242/44.

N-PSDB; AA014410.

New viral proteins from non A-non-B hepatitis agent - used to

treat and prevent enterically-transmitted non-A non-B hepatitis

virus

Claim 3; Page 9; 117pp; English.

A positive clone ETL.1 was identified in a library prepared from

bile of cynomolgus monkeys infected with the Burma strain of ET-NANB.

Both strands of ETL.1 were sequenced. One was designated the

"forward" strand because of statistical similarities to known

proteins and because the forward sequence is known to be

predominantly protein-encoding. Of the three possible reading frames

only the first (ORF 1) is uninterrupted; the other two include

many termination codons. See AA14616 and AA14617.

Sequence 431 AA;

Query Match 22.8%; Score 2057; DB 12; Length 431;

Best Local Similarity 88.4%; Pred. No. 2.8e-159; Matches 381; Conservative 26; Mismatches 24; Indels 0; Gaps 0;

PF 25-JUN-1994; 94US-0279823.
XX
XX 05-APR-1991; 91US-0681078.
PR 17-JUN-1988; 88US-0208997.
PR 11-APR-1989; 89US-0336672.
PR 19-JUN-1989; 89US-0367486.
PR 13-OCT-1989; 89US-0420921.
PR 05-APR-1990; 90US-0505888.
PR 25-JUL-1994; 94US-0279823.
XX

PA (GENE-) GENELABS TECHNOLOGIES INC.

PI Bradley DW, Fry KE, Krawczynski KZ, Reyes GR, Tam A,
PI Yarbough PO;
XX

DR WPI: 1998-446186/38.
DR N-PSDB; AAV54728.
XX

XX Hepatitis E virus DNA - useful for e.g. virus detection and viral
PT protein production
XX

PS Example 3; Columns 43-46; 45pp; English.
XX

CC The present sequence is encoded by the 1.33 kb EcoRI insert of clone
CC ET1.1 of an enterically transmitted nona/nonb viral hepatitis agent
CC (ET-NANB). The sequence in plasmid pTZKFI(ET1.1) carried in E. coli
CC strain B4 is deposited under ATCC 67717. The nucleic acid sequence
CC may be used for identifying and sequencing the entire viral agent
CC (also referred to as HEV), detecting ET-NANB in infected samples,
CC e.g. by specific amplification of virus-derived DNA sequences
CC and for producing recombinant viral proteins for use in vaccines.
XX

XX Sequence. 431 AA;
SQ

Query Match 22.8%; Score 2057; DB 19; Length 431;
Best Local Similarity 88.4%; Pred. No. 2.8e-159;
Matches 381; Conservative 26; Mismatches 24; Indels 0; Gaps 0;

OY 1257 RPAPVAAVIPCPPELEQGLYMPQELTSDVSLVPELTDIVHCRMAAPSQRAVLSTLVG 1316
DB 1 rpvpaavipcppeleqgllympqeltcdsvvlteltdivhcrmaapsqrakavistlvq 60
OY 1317 RYGRRTKLYEAHSDRESLARFPTIGPVQATTCELYELVAMVEKGQDSAVLELDLC 1376
DB 61 rygrrtklynashdreslarfipagvyttceljelveamvekgdgsavleldlc 120
OY 1377 NRDVSRITFFQKXCNKFTTGETIAGKVGQGISAMSKTFCALFGPWRRAIEKEILALLPP 1436
DB 121 nrdsrirtffqkcnkfttgetiagkvgqgisamsktfcalfgpwrariekailallpq 180
OY 1437 NIFYGDATYESFMAAVSAGSCMFENDFSEFDSQNNFSLGLECVMEECGMPQMLIR 1496
DB 181 nifygdafdyfisaavaakasmwyfendfsefdsqnnfslglecavmeecgmpqmlir 240
OY 1497 LYLHVRSAWITLQAPKESLGFPMKHSSEPTLLMNTVMNMAIIAHCEPRDFRVAFAFGD 1556
DB 241 lylhvsawitlqapkeslgtfpmkhsseptllmntvmnmaiiahceprdfrrvaafagd 300
OY 1557 DSVLCSDFYROSNAALTAGGLKLKVDYRPIGLYAGVVAAGLGLTDPVVRFAGRlse 1616
DB 301 dsvlcsdfyrosnaaltagglklkvdyrpigliyagvvvaaglgltldpvyvrfagrllse 360
OY 1617 KMWGPPEPRAEQLACDFLGLTINVAQCVVSVRYGVSPGLVHNLIQMLQTIADGK 1676
DB 361 kmwgppeperaegllrlavsdflrltnvaqmcvdsrvygvspglvhnliqmlqivadgk 420
OY 1677 AHTEETIKPVL 1687
DB 421 ahteesvxpvl 431

Search completed: March 7, 2002, 14:03:47
Job time: 110 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2002, 14:01:57 ; Search time 19.44 Seconds
(without alignments)
1965.565 Million cell updates/sec

Title: US-09-468-147-91

Perfect score: 9013

Sequence: 1 PGITTAIEQAALAAANSALA.....FTETIKPYLDLTNSIQKVE 1698

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued Patents AA:*
2: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7448	82.6	1693	3	US-08-478-507-7
2	7448	82.6	1693	3	US-09-128-275A-7
3	7439	82.5	1693	4	US-08-840-316-1
4	7439	82.5	1693	4	US-08-809-523-1
5	7439	82.5	1693	4	US-08-471-971-1
6	7439	82.5	1693	5	PCR-US93-08849A-1
7	7439	82.5	1693	5	PCR-US93-08849A-1
8	2057	22.8	431	3	US-08-478-507-2
9	2057	22.8	431	4	US-09-128-275A-2
10	515	3.7	1704	4	US-08-485-355B-40
11	374.5	4.2	2205	1	US-08-093-453B-2
12	365	4.0	78	2	US-07-876-941A-31
13	360	4.0	1233	4	US-09-194-613-5
14	233	2.6	2500	3	US-08-801-263A-2
15	233	2.6	2500	3	US-09-102-248-2
16	218	2.4	2512	2	US-08-801-263A-9
17	218	2.4	2512	2	US-09-102-248-9
18	212.5	2.4	2517	2	US-08-801-263A-5
19	212.5	2.4	2517	2	US-09-102-248-5
20	211.5	2.3	2161	3	US-09-081-320-3
21	200	2.2	2431	1	US-07-920-281C-2
22	200	2.2	2431	4	US-08-466-277-2
23	184	2.0	1390	4	US-08-770-544-2
24	178	2.0	34	2	US-07-876-941A-32
25	165.5	1.8	255	3	US-09-081-320-14
26	162	1.8	1456	1	US-08-803-973-2
27	162	1.8	1456	1	US-08-803-973-2

28	158.5	1.8	673	4	US-09-196-387-8	Sequence 8, Appl
29	158.5	1.8	949	4	US-09-196-387-10	Sequence 10, Appl
30	158.5	1.8	1327	4	US-09-196-387-2	Sequence 2, Appl
31	155	1.7	1185	4	US-09-041-886-23	Sequence 23, Appl
32	152.5	1.7	459	4	US-09-080-983-5	Sequence 5, Appl
33	148.5	1.6	3567	2	US-07-642-734C-4	Sequence 4, Appl
34	148.5	1.6	3567	2	US-08-439-009A-4	Sequence 4, Appl
35	145	1.6	1463	1	US-08-157-005-3	Sequence 3, Appl
36	145	1.6	1463	4	US-08-747-863-3	Sequence 3, Appl
37	145	1.6	3724	2	US-08-804-227C-10	Sequence 10, Appl
38	145	1.6	3724	2	US-08-804-198-4	Sequence 4, Appl
39	143.5	1.6	1315	3	US-08-899-595-3	Sequence 3, Appl
40	140.5	1.6	1255	2	US-09-080-897-4	Sequence 4, Appl
41	140.5	1.6	1255	3	US-08-899-595-1	Sequence 1, Appl
42	140.5	1.6	1255	4	US-09-323-735-4	Sequence 4, Appl
43	140	1.6	2152	4	US-09-036-987A-3	Sequence 3, Appl
44	140	1.6	2152	4	US-09-370-700-3	Sequence 3, Appl
45	139	1.5	902	1	US-08-396-479B-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-08-478-507-7
; Sequence 7, Application US/08478507
; Patent No. 6120988
; GENERAL INFORMATION:
; APPLICANT: Reyes, Gregory R
; APPLICANT: Yarbrough, Patricia O
; APPLICANT: Bradley, Daniel W
; APPLICANT: Kracyszynski, Krzysztof Z
; APPLICANT: Tam, Albert
; APPLICANT: Fry, Kirk E
; TITLE OF INVENTION: DNA Sequences of Enterically Transmitted
; TITLE OF INVENTION: No. 6120988-A/No. 6120988-B Hepatitis Viral Agent
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/478,507
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/279,823
; FILING DATE: 25-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/681,078
; FILING DATE: 05-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/505,888
; FILING DATE: 05-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/420,921
; FILING DATE: 13-OCT-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/367,486
; FILING DATE: 16-JUN-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/336,672
; FILING DATE: 11-APR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/208,997
; FILING DATE: 17-JUN-1988

ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0183, 22
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 324-0880
TELEFAX: (650) 324-0960
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1693 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-478-507-7

Query Match 82.6%; Score 7448; DB 3; Length 1693;
Best Local Similarity 81.8%; Pred. No. 0;
Matches 1396; Conservative 110; Mismatches 169; Indels 32; Gaps 6;

QY 1 PGTTAIEQALAAANSALANAVVPRFLSRVOTELLINLMOPROLVFRPEVLMNPIOR 60
DB 10 PGTTAIEQALAAANSALANAVVPRFLSHOOIELINLMOPROLVFRPEVLMNPIOR 69
QY 61 VIHNELOYCARAGRCLEVGAPRSINDPNVLRHRCFLRPYGRDQRYMSAPTRGPAN 120
DB 70 VIHNELEYCARAGRCLEIGAPRSINDPNVLRHRCFLRPYGRDQRYMSAPTRGPAN 129
QY 121 CRRSALRGPRADRYTCFEGFSKCAFAAETGVALYSLHLMRPADVAEAMRHGXTLYAA 180
DB 130 CRRSALRGPRADRYTCFEGFSKCAFAAETGVALYSLHLMRPADVAEAMRHGXTLYAA 189
QY 181 LHLPEVLLPGYHTTSTYLLIHDGDAVVEGDTSGYNDVSLIRAMIRTKTVGH 240
DB 190 LHLPEVLLPGYHTTSTYLLIHDGDAVVEGDTSGYNDVSLIRAMIRTKTVGH 249
QY 241 PLVIERVRAIGHFVLLTAABEPSMPVYPRSTEVYVRSIFGPGSPSLFPSCSTR 300
DB 250 PLVIERVRAIGHFVLLTAABEPSMPVYPRSTEVYVRSIFGPGSPSLFPSCSTR 309
QY 301 STEFHAVPIIMDRIMLFGATLDDQACCSRMTYLRGISTKYTVGALVANECHNASEDL 360
DB 310 STEFHAVPIIMDRIMLFGATLDDQACCSRMTYLRGISTKYTVGALVANECHNASEDL 369
QY 361 TAVITAAVLTICHQRIKRLQAIKSKMRRLGVEHAOKETIRLXSWLEKSGROYIPROLO 420
DB 370 TAVITAAVLTICHQRIKRLQAIKSKMRRLGVEHAOKETIRLXSWLEKSGROYIPROLO 429
QY 421 FYAGCRWLASGFHLDPRLVFEDESVPKCRKTEFLKRVAGKFCCEMRLQOECTCLEPAE 480
DB 430 FYAGCRWLASGFHLDPRLVFEDESVPKCRKTEFLKRVAGKFCCEMRLQOECTCLEPAE 489
QY 481 GLVGHGHNDNEAYEGSEVDPAEPALHDVSGTYAHHGHOELALYRALNVQDIAAARSKLT 540
DB 490 GLVGHGHNDNEAYEGSEVDPAEPALHDVSGTYAHHGHOELALYRALNVQDIAAARSKLT 549
QY 541 ATTELVASDRLCERVTGLNKTFTTVDGAHLEANGPQOYVILSPDASQSGAGSHSLT 600
DB 550 ATTELVASDRLCERVTGLNKTFTTVDGAHLEANGPQOYVILSPDASQSGAGSHSLT 609
QY 601 YELTPAGLQVIRISSNGLDCTATFPFGAGAPSAAPGEVAACALYRYNRPTQHSLSLGLAM 660
DB 610 YELTPAGLQVIRISSNGLDCTATFPFGAGAPSAAPGEVAACALYRYNRPTQHSLSLGLAM 669
QY 661 LHPBELLGIFPPSPGHIMESANPCGEGTLYTRTWS-TSGSSSDSP-----PEAADA 714
DB 670 LHPBELLGIFPPSPGHIMESANPCGEGTLYTRTWS-TSGSSSDSP-----PEAADA 729
QY 715 MAATGDLHPSTPPVDIWLPPRSEFOVDAAVY-PPAPDPAGLP-GPVYVLLPPPPPVH 772
DB 730 RAATPTL-----AAPLPPADPSPSPAPLADPAGGATAG 766
QY 773 KRISIP-PPSRNRLLTYTPDGAKEYAGSLFESDCDMLVANSNPGRPGGLCHAFYQRP 831

DB 767 APAITHQTAHRRLFTYPDGSKVFAAGSLFESTCTMLVANSNDHRRGGGLCHAFYQRP 826
QY 832 EAFYPTPEFTIMREGLAATYTLPRPIIAVADRYEONPKLEAYETCSRKRTAAYPLL 891
DB 827 ASFDAASFWRDGAAYATTLPRPIIAVADRYEONPKLEAYETCSRKRTAAYPLL 886
QY 892 GSGIYQVPSLSDAMEERNHRPDELYTEPAAWFEAKNPPAPVLTJEDTARTANLAL 951
DB 887 GTGIVQVPIGSPFDAMERNHRPDELYTEPAAWFEAKNPPAPVLTJEDTARTANLAL 946
QY 952 EIDAAETVRACAGCTISPGIYHVOPTAGVPSGSKRSIQOGDVVVVYPTRELRMSWR 1011
DB 947 EIDAAETVRACAGCTISPGIYHVOPTAGVPSGSKRSIQOGDVVVVYPTRELRMSWR 1006
QY 1012 RGFPAFTPHTAARVYTGRRVYIDEAPSLPHLLLMORASVHLLGDNPQIPADIFEHA 1071
DB 1007 RGFPAFTPHTAARVYTGRRVYIDEAPSLPHLLLMORASVHLLGDNPQIPADIFEHA 1066
QY 1072 GLVPATIRPELAPTSMMXVTHRCPADVCYLIRGAYPKIOTTSRYLRSLFNNPRAIGOKLY 1131
DB 1067 GLVPATIRPELAPTSMMXVTHRCPADVCYLIRGAYPKIOTTSRYLRSLFNNPRAIGOKLY 1126
QY 1132 TOAKANPAGATVHBAOGATFETTTIATADARGLIOSSRAHAIVALTRETEKCVILDA 1191
DB 1127 TOAKANPAGATVHBAOGATFETTTIATADARGLIOSSRAHAIVALTRETEKCVILDA 1186
QY 1192 PGLIREVIGSIDVYVNNFVLAGEVGHRSVYPRGPNQDNLGLOAFPPSCQISATHQLA 1251
DB 1187 PGLIREVIGSIDVYVNNFVLAGEVGHRSVYPRGPNQDNLGLOAFPPSCQISATHQLA 1246
QY 1252 EELGHRPAPVAAVLPCCPELOGLLYMPOELVYSDSVLFEELTDIYHCHMAAPSOXKAYL 1311
DB 1247 EELGHRPAPVAAVLPCCPELOGLLYMPOELVYSDSVLFEELTDIYHCHMAAPSOXKAYL 1306
QY 1307 SELVGRYGGRTKLYNASHSDVDRSLARFIPAIGPVQVTCCELYELVEAMVEKQDGSAYL 1366
QY 1372 ELDLCNRDVSRTTFQKXCNKFTTGETTAHGVGQGISAMSKTFCLFEPWRAIRKETL 1431
DB 1367 ELDLCNRDVSRTTFQKXCNKFTTGETTAHGVGQGISAMSKTFCLFEPWRAIRKETL 1426
QY 1432 ALLPNIIFYGDVYEEESVFAAASGACSCWPFENDSEFPSTONNFSGLGECVMECEGMP 1491
DB 1427 ALLPNIIFYGDVYEEESVFAAASGACSCWPFENDSEFPSTONNFSGLGECVMECEGMP 1486
QY 1492 QMLIRLYHLVRSAMILLQAPKESILKGFWRKHSEPGTLLMNTVWMAITIAHCEFRDFRYA 1551
DB 1487 QMLIRLYHLVRSAMILLQAPKESILKGFWRKHSEPGTLLMNTVWMAITIAHCEFRDFRYA 1546
QY 1552 ARKGDSVYLCDYRQSRNAALLIAGCGLKLVDYRPIELLYAGVYVAPRLGLTLPDVYRA 1611
DB 1547 ARKGDSVYLCDYRQSRNAALLIAGCGLKLVDYRPIELLYAGVYVAPRLGLTLPDVYRA 1606
QY 1612 GRLSEKMGCPGERAEOLALVACDFLRGLTNAOVVDVYVSRVYGVSPGLVNLJGMLDT 1671
DB 1607 GRLSEKMGCPGERAEOLALVACDFLRGLTNAOVVDVYVSRVYGVSPGLVNLJGMLDT 1666
QY 1672 IADGKAHFTETIKPVLDLTNSIIORVE 1698
DB 1667 IADGKAHFTETIKPVLDLTNSIIORVE 1693

RESULT 2
US-09-128-275A-7
Sequence 7, Application US/09128275A
Patent No. 6229005
GENERAL INFORMATION:
APPLICANT: Reyes, Gregory R
APPLICANT: Yarbough, Patrice O
APPLICANT: Bradley, Daniel W
APPLICANT: Krzewczynski, Krzysztof Z

APPLICANT: Tam, Albert
APPLICANT: Fry, Kirk E
TITLE OF INVENTION: DNA Sequences of Enterically Transmitted
TITLE OF INVENTION: No. 6229005-A/No. 6229005-B Hepatitis Viral Agent
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSER: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/128,275A
FILING DATE: 03-AUG-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/279,823
FILING DATE: 25-JUL-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/681,078
FILING DATE: 05-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/505,888
FILING DATE: 05-APR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/420,921
FILING DATE: 13-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/367,486
FILING DATE: 16-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/336,672
FILING DATE: 11-APR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/208,997
FILING DATE: 17-JUN-1988
ATTORNEY/AGENT INFORMATION:
NAME: Petithory, Joanne R
REGISTRATION NUMBER: 42,995
REFERENCE/DOCKET NUMBER: 4600-0183.24
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 324-0880
TELEFAX: (650) 324-0960
INFORMATION FOR SEQ. ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1693 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-128-275A-7

Query Match 82.6%; Score 7448; DB 4; Length 1693;
Best Local Similarity 81.8%; Pred. No. 0;
Matches 1396; Conservative 110; Mismatches .169; Indels 32; Gaps 6;

QY 1 PGTTTAEQAALAAANSALANAVVRRPFLSRVOTELLINIMOPRQOLVREPVLMNHPIOR 60
DB 10 PGTTTAEQAALAAANSALANAVVRRPFLSRVOTELLINIMOPRQOLVREPVLMNHPIOR 69
QY 61 VINHELEOYARARAGRCLEVGARHRSINDNPNVYHRCFLRPYGDVORWYSAPTRGPAAN 120
DB 70 VINHELEOYARARAGRCLEVGARHRSINDNPNVYHRCFLRPYGDVORWYSAPTRGPAAN 129
QY 121 CRSSALGLPADRTYCFDFSRCAFAETGVALYSLHDLMPADYADAMAHGXTRLYAA 180
DB 130 CRSSALGLPADRTYCFDFSRCAFAETGVALYSLHDLMPADYADAMAHGXTRLYAA 189

QY 181 LHLPEVLLPPGYHTHTSYLLIHGDRAVVTYEDGTSGAGNHDSILRAMIRTKYIGDH 240
DB 190 LHLPEVLLPPGYHTHTSYLLIHGDRAVVTYEDGTSGAGNHDSILRAMIRTKYIGDH 249
QY 241 PLVIERVRAIGCHEVLLTLAAPESPMPVYVPSTREVVYRSJFGPGSPSLPSPASCTK 300
DB 250 PLVIERVRAIGCHEVLLTLAAPESPMPVYVPSTREVVYRSJFGPGSPSLPSPASCTK 309
QY 301 STEHAVVHIDRLMLFGALLDQAFCCSRMLTYLRISIKYVYGAIVANEGMASDAL 360
DB 310 STEHAVVHIDRLMLFGALLDQAFCCSRMLTYLRISIKYVYGAIVANEGMASDAL 369
QY 361 TAYTTAAYLTTCORRYLRTQAIKGMRLGVEHROKFTRLYSWLFPSKGRDYPGQOLO 420
DB 370 TAYTTAAYLTTCORRYLRTQAIKGMRLGVEHROKFTRLYSWLFPSKGRDYPGQOLO 429
QY 421 FYACRRMLSAGFHLDPRLVLFDESVPQRCTFLKRYAGRCCPMWLGOCCTCELPAPAE 480
DB 430 FYACRRMLSAGFHLDPRLVLFDESVPQRCTFLKRYAGRCCPMWLGOCCTCELPAPAE 489
QY 481 GLVGDHGDNEAYEGSEVDPAEPALHDVSGTYAVHGHOLEALRYALVPODIAARSLT 540
DB 490 GAVGDQCHDNEAYEGSEVDPAESAISDISGYYVPTALPLQALDPAETVARAGRLT 549
QY 541 ATVELVASPDLRECRVTGKTFRTYVDGAHLEANGPEQYVLSFDSRSMGASHSLT 600
DB 550 ATVELVASPDLRECRVTGKTFRTYVDGAHLEANGPEQYVLSFDSRSMGASHSLT 609
QY 601 YELTPALQVIRISSNGLDCTATPPGAPSAPEAEVAFSCALRYRRTQSHSLTGLW 660
DB 610 YASAPALEVRYVAAAGLDHRAVFAVSPRSAPBEVAFSCALRYRRTQSHSLTGLW 669
QY 661 LHPEGLIGTPEPSPGHIMESANPFCGEGILYRTWS-TSGFSSDESP-PEAAPA 714
DB 670 FHPEGLIGTPEPSPGHIMESANPFCGEGILYRTWS-EVAPAPDLGFMSEPSIS 729
QY 715 MAATPGLPHSTPPVSDTWLPRPSEEFQVDAAPV-PPAPDAGLP-GPVYLTTPPPPVH 772
DB 730 RAATPPL-AA 766
QY 773 KPSIP-PPSRNRLLLYTPDGAKYVAGSLFESDCCDWLVNANSPGHRPGGICHAIFYORF 831
DB 767 APATHTQARHRLLFYTPDGSKYFAGSLFESTCTWLVNANSPGHRPGGICHAIFYORF 826
QY 832 EAFYPTFEIMEGLAAYTLPRPIIHAVADYRVEQNPRLAAYRETCSRGAAYPL 891
DB 827 ASFDASEFVMDGAAYTLPRPIIHAVADYRVEQNPRLAAYRETCSRGAAYPL 886
QY 892 GSGIYQVPSLSFDAMERNHRPGDELYLLEPAAMFEANRPAQVYLITETARTANLAL 951
DB 887 GTGIYQVPIGSPFAMERNHRPGDELYLLEPAAMFEANRPAQVYLITETARTANLAI 946
QY 952 EIDATEVGRACACCTISPGIVHYQTAGVYSGSKSSIOGGDVYVVPRELNRSMR 1011
DB 947 ELDSATDVGRACACCTISPGIVHYQTAGVYSGSKSSIOGGDVYVVPRELNRSMR 1006
QY 1012 RGFATPHTAARTYIGRRVVIDEAPSLPRLLLHMQRASSVHLGDPNOIPAIIDEHA 1071
DB 1007 RGFATPHTAARTYIGRRVVIDEAPSLPRLLLHMQRASSVHLGDPNOIPAIIDEHA 1066
QY 1072 GLVPAIRPELAPYSWXYTHRCPADVCELRGAYPKIQTTSRVLSLFWNEPAIGOKLYX 1131
DB 1067 GLVPAIRPELAPYSWXYTHRCPADVCELRGAYPKIQTTSRVLSLFWNEPAIGOKLYX 1126
QY 1132 TOAKKANPGAITVHEAGATFTETITIAATDAGLIQSSRAHAIVALTRTECVILDA 1191
DB 1127 TOAKKANPGAITVHEAGATFTETITIAATDAGLIQSSRAHAIVALTRTECVILDA 1186
QY 1192 PGLIREVIGSDVIVNNEFLAAGEVGHARPSVYPRGNPDONLGLQAPPPSCQISAYHOLA 1251
DB 1187 PGLIREVIGSDVIVNNEFLAAGEVGHARPSVYPRGNPDONLGLQAPPPSCQISAYHOLA 1246
QY 1252 EELGHRPAPVAAVLPPECELEGGILYMPQELTYSDSVLYVELTIDIVHCRAAPASQRAVL 1311

```

Db 1247 EELGHRPVPVAAVLPCPELEQGLTLXLPQELTTCDSVYTFELTDIVHCRMAAPSQRVAVL
QY 1312 STLVGRYGRRTKLYEAHSHVRESLARFPITPTGVOATTCETELVEIEMAWEKGDSSAVL
Db 1307 STLVGRYGRRTKLYEAHSHVRESLARFPITPTGVOATTCETELVEIEMAWEKGDSSAVL
QY 1372 ELDLCNRDVSRTTFQKXCNKFTTGETIAHGKVGQGISAMSKTFCALFGFWFAIRKEIL
Db 1367 ELDLCNRDVSRTTFQKXCNKFTTGETIAHGKVGQGISAMSKTFCALFGFWFAIRKEIL
QY 1432 ALPPPIIFGDAAVEEVPAAVSGACSCWFEFDESEFSTONNFSLGLECVMEECGMP
Db 1427 ALPPQVVFEGDADDTVFSAVAARAAKASWFEFDESEFSTONNFSLGLECVMEECGMP
QY 1492 QMLIRLYHIVRSAMIIQAPESLKGFKHSGEPGLTNVNMMAIIAHCEYEPDRVA
Db 1487 QMLIRLYHIVRSAMIIQAPESLKGFKHSGEPGLTNVNMMAIIAHCEYEPDRVA
QY 1552 AFKGDSDVYLCSDYRQSRNAAALIAAGCGLKLVYDRPIGLYAGVVAAPGLTLPDVVRA
Db 1547 AFKGDSDVYLCSDYRQSRNAAALIAAGCGLKLVYDRPIGLYAGVVAAPGLTLPDVVRA
QY 1612 GRLESENMKGPPRRAQLRAVCDPLRLGLTNVAQVCVDVYSRYGVSPGLVHNLIGMLQT
Db 1607 GRLESENMKGPPRRAQLRAVCDPLRLGLTNVAQVCVDVYSRYGVSPGLVHNLIGMLQT
QY 1672 IADGRAHFTETIKVLDLTFNSIOTRYE 1698
Db 1667 VADGRAHFTESKVPVLDLTFNSILCRVE 1693

```

RESULT 3

US-08-840-316-1

Sequence 1, Application US/08840316

Patent No. 6054567

GENERAL INFORMATION:

APPLICANT: Emerson, Suzanne U., Purcell, Robert H.,

APPLICANT: Tsarev, Sergei, A., and Robinson, Robin A.

TITLE OF INVENTION: Recombinant Proteins Of

TITLE OF INVENTION: A Pakistani Strain Of Hepatitis E And Their

TITLE OF INVENTION: Use In Diagnostic Methods And Vaccines

NUMBER OF SEQUENCES: 111

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN

STREET: 345 PARK AVENUE

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: USA

ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WORDPERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/840,316

FILING DATE: 11-APR-1997

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Richard W. Bork

REGISTRATION NUMBER: 36,459

REFERENCE/DOCKET NUMBER: 2026-4255

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 758-4800

TELEFAX: (212) 751-6849

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1693 AMINO ACID RESIDUES

```

; TYPE: AMINO ACID
; STRANDEDNESS: UNKNOWN
; TOPOLOGY: UNKNOWN
US-08-840-316-1

```

Query Match 82.5%; Score 7439; DB 3; Length 1693;
Best Local Similarity 81.6%; Pred. No. 0;
Matches 1397; Conservative 111; Mismatches 162; Indels 42; Gaps 7;

```

QY 1 PGITTAIEQAALAAANSALANAVVPRFLSRVQTELLINIMOROLVFRREVLMNPIOR 60
Db 10 PGITTAIEQAALAAANSALANAVVPRFLSRVQTELLINIMOROLVFRREVLMNPIOR 69
QY 61 VINHELEIYCRARAGCLEVGAHPRSIINDPNVLAHCFELRPVORDVQRMVSAFTPGPAAN 120
Db 70 VINHELEIYCRARAGCLEVGAHPRSIINDPNVLAHCFELRPVORDVQRMVSAFTPGPAAN 129
QY 121 CRRSALRGLEPPADRTYCFDGFSCAPAEETGVALYSLHDMIPADVAEAMARHGXYRLYAA 180
Db 130 CRRSALRGLEPPADRTYCFDGFSCAPAEETGVALYSLHDMIPADVAEAMARHGXYRLYAA 189
QY 181 LHLPREVLLPFGTYHTTSYLLIHDGRVAVTYEGDTSAGYNHDSVILRAIRTKYVGDH 240
Db 190 LHLPREVLLPFGTYHTTSYLLIHDGRVAVTYEGDTSAGYNHDSVILRAIRTKYVGDH 249
QY 241 PLVIERVRAIGCHFVLLTLAAPERSPMPYVPRSTEVYRSIFGPGSPSLFPASCSTK 300
Db 250 PLVIERVRAIGCHFVLLTLAAPERSPMPYVPRSTEVYRSIFGPGSPSLFPASCSTK 309
QY 301 STEFAVAVVHIWDRMLFGATLDDQAFCCSRMLTYLRGISYKVTGVALVANEGMNASDAL 360
Db 310 STEFAVAVVHIWDRMLFGATLDDQAFCCSRMLTYLRGISYKVTGVALVANEGMNASDAL 369
QY 361 TAITTAAYLTICQRLRQATISKGMRLGVHAQCEITRLYSWLEKSGRDYIPGRQ 420
Db 370 TAITTAAYLTICQRLRQATISKGMRLGVHAQCEITRLYSWLEKSGRDYIPGRQ 429
QY 421 FYACRRMLSAGHILPRVYLFEDSEVPCRCRTFLKLVAGFCFCFMMLQCECTFLEPAE 480
Db 430 FYACRRMLSAGHILPRVYLFEDSEVPCRCRTFLKLVAGFCFCFMMLQCECTFLEPAE 489
QY 481 GLVGDHGHNEAVEGSEVDPAEPALDVSGTYAVHGHQLEALYALNVPDIAARASRLT 540
Db 490 GLVGDHGHNEAVEGSEVDPAEPALDVSGTYAVHGHQLEALYALNVPDIAARASRLT 549
QY 541 ATVEIYASPDRLCEKRVNLGNKTRTIVVQGAHLEANGPQGVYVSPASQSMKAGSHSLT 600
Db 550 ATVEIYASPDRLCEKRVNLGNKTRTIVVQGAHLEANGPQGVYVSPASQSMKAGSHSLT 609
QY 601 YELTPAGLOVRISNGLDCTATPPPGGAPSAAPGEVAAPFCASALYRYNRFTORSLTGLY 660
Db 610 YELTPAGLOVRISNGLDCTATPPPGGAPSAAPGEVAAPFCASALYRYNRFTORSLTGLY 669
QY 661 LHPGGLGIFPPSPGHIMESANPFCGEGTLYTRWS-----TSFGSSDFSPP 709
Db 670 LHPGGLGIFPPSPGHIMESANPFCGEGTLYTRWS-----TSFGSSDFSPP 724
QY 710 AAAPMAATPGLPHSTPPVPSDITVLPPESEFOVDAAPV-PPAPDPA-GLPGVYVLTTPP 767
Db 725 PSIPSRATP-----TP-----AAAPPPAPDPSPTLSAPARGEPP 761
QY 768 PPVHKKPSIP-PPSRNRRLTYTPDQAKYVAGSLFSSDDMLVNASNPGHPPGGI-CHAF 826
Db 762 GATARAAPAITTHQTAHRLRLFTTPDQSKYFAGSLFESTCTWLNASNVDRHPPGGI-CHAF 821
QY 827 YORPEAFYPTFEINREGLAAYTLTPRPIIHAVADRYEONKRLAEAYRETCSRRTGA 886
Db 822 YORPEAFYPTFEINREGLAAYTLTPRPIIHAVADRYEONKRLAEAYRETCSRRTGA 881
QY 887 ATPPLGSGIYQVAVSLSPDAMEKRNHRGDELVLTPEPAAMFEANKPAQPVLTITTEDYART 946
Db 882 ATPPLGSGIYQVAVSLSPDAMEKRNHRGDELVLTPEPAAMFEANKPAQPVLTITTEDYART 941

```



```

Db 550 ATWKSQVNDGRIDCELTLLNKTFRTSFVDAVLETENGPERHNLSEFASOSTMAAGFSILT 609
QY 601 YELTRAGIQVRRSSNGDCTATFPFGCAPSAAPGEVAACSLXRXNRRTQHSILGLGIM 660
Db 610 YASASAGLEVRYVAAGLDHRAVAPGVSPRSAPGEVTAFCASALYRNRRAQRLSTLGNW 669
QY 661 LHPEGLIGTFPPSPGHMESANPFCEGELTYRTWS-----TSGFSDSPPE 709
Db 670 FHPEGLIGTFAPSPGHVESANPFCEGESTLYTRTWSEVDAPSPAPDGLGFTS-----E 724
QY 710 AAAPAMATPGLPHSTPPVSDIWLPPRSEEFQVDAAPV-PPAPDPA-GLPGPVLTTPP 767
Db 725 PSIPSAAP-----TP-----AAPLPAPDPSPPLSAPAREPAP 761
QY 768 PPPVHKPSIP-PSNRRLLYTPPGAKYAGSLFESDCLWVMSNPCHRGCGGCHAF 826
Db 762 GATAPAPATTHOTAHRRLLFTYPPDSKVFAGSLFESTCTWLVNSNVDRHGGGLCHAF 821
QY 827 YORPEAPYPTFEIMREGIAAYTLPPRLTHAVADYRVEONPKRLAAVRETCSRGT 886
Db 822 YORYPASFPDASFVMDGAAAYTLPPRLTHAVADYRLEHNPKRLAAVRETCSRGT 881
QY 887 AYPPLGSGIYQVPSLSPDAMERNHRPGDELYLTERPANMEANKPAQVLTITEDTANT 946
Db 882 AYPPLGTGIYQVPIGSPDAMERNHRPGDELYLTERPANMEANKPAQVLTITEDTANT 941
QY 947 ANLALEIDAAETVGRACACTISPGIVHQTAGVPGSGKSRSTOOGDQDVVVVPPREAR 1006
Db 942 ANLALEIDAAETVGRACACTISPGIVHQTAGVPGSGKSRSTOQDVVVVPPREAR 1001
QY 1007 NSMRRRGEAFTPHFAARTYIGRRVVIDEAPSLPRLHLLHMQRASSVHLGDPPNIPAI 1066
Db 1002 NAMRRRGFAFTPHFAARTYIGRRVVIDEAPSLPRLHLLHMQRATVHLGDPPNIPAI 1061
QY 1067 DEEHAGLVPAIRPELAPTSWMXVYTHRCPADVCELLRGAYPKIQTTSRVLRSLFWNEPAIG 1126
Db 1062 DEEHAGLVPAIRPELAPTSWMXVYTHRCPADVCELLRGAYPKIQTTSRVLRSLFWNEPAIG 1121
QY 1127 OKLVYTOAKKANPGAITHEHOGATFETTTIATADAGLTIQSSAAHIVALTTRTEKC 1186
Db 1122 OKLVYTOAKKANPGAITHEHOGATFETTTIATADAGLTIQSSAAHIVALTTRTEKC 1181
QY 1187 VILDAAGLREYISDVIVNNEFLAGGEVXHRPSVYIPRGNPDONLGTLOAFPPSCQISA 1246
Db 1182 VILDAAGLREYISDVIVNNEFLAGGEVXHRPSVYIPRGNPDONLGTLOAFPPSCQISA 1241
QY 1247 YHQLAEELGHRAPVAAYLPPCPELEOGLLYMPQELTVSDSVLVEFLTDIVHCRMAAPSO 1306
Db 1242 FHELEAEELGHRAPVAAYLPPCPELEOGLLYMPQELTVSDSVLVEFLTDIVHCRMAAPSO 1301
QY 1307 RKAIVSTIVGRYGRPKLYLEAASDVRESLARPIPTIGVOATTCGLYLVEAMVEKGD 1366
Db 1302 RKAIVSTIVGRYGRPKLYLEAASDVRESLARPIPTIGVOATTCGLYLVEAMVEKGD 1361
QY 1367 GSAVLELDLCNDVSRITFFQCKXCKFTTGETIAGHKVQGISAMSKTECALFGPWFRAI 1426
Db 1362 GSAVLELDLCNDVSRITFFQCKXCKFTTGETIAGHKVQGISAMSKTECALFGPWFRAI 1421
QY 1427 EKEIALLPPNIFYGDAAVEEYFAAASGAGSCWFEENDSFSDSTONNLSGLECVNE 1486
Db 1422 EKEIALLPPNIFYGDAAVEEYFAAASGAGSCWFEENDSFSDSTONNLSGLECVNE 1481
QY 1487 ECGMOMLIRLYHLVRSANITLAPKESLKGFKKHSGEPTLLMNTVMMAITLHACYER 1546
Db 1482 ECGMOMLIRLYHLVRSANITLAPKESLKGFKKHSGEPTLLMNTVMMAITLHACYER 1541
QY 1547 DERVAAFKDDSVLCSVDROSRNAAALTAGGLKLVYRIGLYAGVVAAPGLGTLDP 1606
Db 1542 DLQVAAFKDDSVLCSVDROSRNAAALTAGGLKLVYRIGLYAGVVAAPGLGTLDP 1601
QY 1607 VVREAGRLSEKNWGPGERAEQRLAVCDLRLTNVAOCVVDVYSVRYVSGPLVHNL 1666

```

```

Db 1602 VVREAGRLSEKNWGPGERAEQRLAVCDLRLTNVAOCVVDVYSVRYVSGPLVHNL 1661
QY 1667 GMLQTIADGKAHPETERIKPVLDTNSIORYE 1698
Db 1662 EMLQAVADGKAHPETERIKPVLDTNSIORYE 1693

RESULT 5
US-08-471-971-1
; Sequence 1, Application US/08471971
; Patent No. 6287759
; GENERAL INFORMATION:
; APPLICANT: Tsarev, Sergei. A., Emerson,
; APPLICANT: Suzanne U., Purcell, Robert H.
; TITLE OF INVENTION: Recombinant Proteins Of
; TITLE OF INVENTION: A Pakistani Strain Of Hepatitis E And Their
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,971
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US07/947,263
; FILING DATE: 18-SEP-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Richard W. Bork
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4032US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1693 AMINO ACID RESIDUES
; TYPE: AMINO ACID
; STRANDEDNESS: UNKNOWN
; TOPOLOGY: UNKNOWN
; US-08-471-971-1

Query Match 82.5%; Score 7439; DB 4; Length 1693;
Best Local Similarity 81.6%; Pred. No. 0;
Matches 1997; Conservative 111; Mismatch 162; Indels 42; Gaps 7;

```


Qy	181	LHPREVLPGVHTTSTYLLIHDOGRAVVTYGDHSAGINHDSYLLRMIITTKYIGDH	240
Db	190	LHPREVLPGVHTRTASYLLIHDOGRAVVTYIGDHSAGINHDSYLLRMIITTKYIGDH	249
Qy	241	PLVIERVATIGCHVLLTLTAAPESPMPVYVYPRSTEVYVRSIFGGSGPSLFPSCSTK	300
Db	250	PLVIERVATIGCHVLLTLTAAPESPMPVYVYPRSTEVYVRSIFGGSGPSLFPSCSTK	309
Qy	301	STFAVYVHTWDRMLMFGALLDOAFCCSMLMTYLGISYKYTVGALVANBESMANSEDAL	360
Db	310	STFAVYVHTWDRMLMFGALLDOAFCCSMLMTYLGISYKYTVGALVANBESMANSEDAL	369
Qy	361	TAXITAAVLTICHOHYRTQOASIKGMARLGEVHAQCFITRLYSWLEKSGRDTGOROLO	420
Db	370	TAVITAAVLTICHOHYRTQOASIKGMARLGEVHAQCFITRLYSWLEKSGRDTGOROLO	429
Qy	421	FYAOCRMLSAGFHLDRVLVDESVCRCRTEFLKRVAGFCCFMWLGQECTCELEPAE	480
Db	430	FYAOCRMLSAGFHLDRVLVDESVCRCRTEFLKRVAGFCCFMWLGQECTCELEPAE	489
Qy	481	GLVGDHGDNDNAYGSSVDPAEPAHNDVSTVAVHNOLEALYRALVNOODIAAASRLT	540
Db	490	GLVGDHGDNDNAYGSSVDPAEPAHNDVSTVAVHNOLEALYRALVNOODIAAASRLT	549
Qy	541	ATVELVASPDRECRVYLGKNTFRTTYVDSGANDHEANGPEDOYVLSFEDASROSAGSHSLT	600
Db	550	ATVELVASPDRECRVYLGKNTFRTTYVDSGANDHEANGPEDOYVLSFEDASROSAGSHSLT	609
Qy	601	YELPRAGLOVRISNGIDLCATATPPRGGAPAAAGEVAALCSALYRNETQHSILTGILM	660
Db	610	YELPRAGLOVRISNGIDLCATATPPRGGAPAAAGEVAALCSALYRNETQHSILTGILM	669
Qy	661	LHPGGLIGIFEPFSGHIMESANPFCGEGLYRTWS-----TSGSPSPSPE	709
Db	670	LHPGGLIGIFEPFSGHIMESANPFCGEGLYRTWS-----TSGSPSPSPE	724
Qy	710	AAAPMAATPGLPHSTPSPVSDIVLPPSEEPVDAAPV-PPAPDPA-GLPGRVYLTPRP	767
Db	725	AAAPMAATPGLPHSTPSPVSDIVLPPSEEPVDAAPV-PPAPDPA-GLPGRVYLTPRP	771
Qy	768	PPVPHKSIP-PPSRRRLLYTYPDGAKKYVAGSLFSSDCMWLVNANPCNHPRGGGICJAF	826
Db	762	GATARAARITQYAKRHRLLFTYPRDGSKYVAGSLFSTCWLNVNANPCNHPRGGGICJAF	821
Qy	827	YOREPEAFYPTFIMREGLAATYLTTPRPIIHAAVPDYRVEONPKRLAAAYRETCSRGT	886
Db	822	YOREPEAFYPTFIMREGLAATYLTTPRPIIHAAVPDYRVEONPKRLAAAYRETCSRGT	881
Qy	887	AYPLGSGITVYRVSLSFDAMERNHRHGBDELYLTERAAMFEANKRAQVYLITTEDTART	946
Db	882	AYPLGSGITVYRVSLSFDAMERNHRHGBDELYLTERAAMFEANKRAQVYLITTEDTART	941
Qy	947	ANLAEIDAATAEVORACAGCTISPGIYHYOFTAGVSGSKRSIOGDVYVVPYPRRLR	1006
Db	942	ANLAEIDAATAEVORACAGCTISPGIYHYOFTAGVSGSKRSIOGDVYVVPYPRRLR	1001
Qy	1007	NSMRRRGPAFTPHTAARVTIGRRVYIDEAPSLPHLLLHMQRASSVHLCDPNOIPAI	1066
Db	1062	NSMRRRGPAFTPHTAARVTIGRRVYIDEAPSLPHLLLHMQRASSVHLCDPNOIPAI	1061
Qy	1127	OKLVYTOAKAANKAGATVVEAGATFETETIATADAGLIQSSRAHAIVALTHTRETK	1186
Db	1122	OKLVYTOAKAANKAGATVVEAGATFETETIATADAGLIQSSRAHAIVALTHTRETK	1181
Qy	1187	VILDAPGLREVGISDVIYVNNFPLAGGEVGHNRSVYPRGNPDONIGTLOAFPPSCQISA	1246
Db	1182	VILDAPGLREVGISDVIYVNNFPLAGGEVGHNRSVYPRGNPDONIGTLOAFPPSCQISA	1241

Qy	1247	YHAAEELGHRHPAAVAALPCPELCSOGLLYMPOETVSDVLAVELDIYHCRMAAPSO	1306
Db	1242	FHEIAEEIGHHPAAVAALPCPELCSOGLLYLPQELITCTDSVTFELDIYHCRMAAPSO	1301
Qy	1307	KKAVLSTLVGRYGRRTALYEAASHDVRESLARIPITIGVQATCELELVEAMEWKEKD	1366
Db	1302	KKAVLSTLVGRYGRRTALYEAASHDVRESLARIPITIGVQATCELELVEAMEWKEKD	1361
Qy	1367	GSAAVELDLCNRDYSRITFFQKKCNKFTTGETIAGHKVGQGISAMSKTFCALFGWFPAT	1426
Db	1362	GSAAVELDLCNRDYSRITFFQKKCNKFTTGETIAGHKVGQGISAMSKTFCALFGWFPAT	1421
Qy	1427	EKETILLPPIIFGDAYEESVFAAASVSGSCMVEENFSEFSDTONNFSGLCEVYE	1486
Db	1422	EKATILLPQGVFEGFADPDTJVSAAVAAAKAMVTEENFSEFSDTONNFSGLCEAIME	1481
Qy	1487	ECGMPOMLIRLYHLYRSAMITQAPKESLGGFWKKHSGEPTLLMTVMNMAIIAHCYEFR	1546
Db	1482	ECGMPOMLIRLYHLYRSAMITQAPKESLGGFWKKHSGEPTLLMTVMNMAIYTHCYDFR	1541
Qy	1547	DFRPAARKGDDSVYVLCSDYRQSRNMAALIAAGCGLKLVYRIRIGLYAGVYAPRGLGTPD	1606
Db	1542	DLYQAAARKGDDSVYVLCSEYRQSGAAVYLLIAGCGLKLVYRIRIGLYAGVYAPRGLGTPD	1601
Qy	1607	VVRFAGRLSEKNMGPGGERAEQRLAVCDLRIGLTIVVAQVCDVYVSRYVGVSPGLVHNL	1666
Db	1602	VVRFAGRLTEKNMGPGGERAEQRLAVSDFRLTIVVAQVCDVYVSRYVGVSPGLVHNL	1661
Qy	1667	GMLOTLADGKAHFTETAKPVLDTNLSIQAVE	1698
Db	1662	EMLOVAVDGAHFTESYKPYVLDTNLSICVE	1693

```

1      RESULT      6
2      PCT-US93-08849A-1
3      ; Sequence 1, Application PC/TUS9308849A
4      ; GENERAL INFORMATION:
5      ; APPLICANT:
6      ; TITLE OF INVENTION: Recombinant Proteins Of
7      ; TITLE OF INVENTION: A Pakistani Strain Of Hepatitis E And Their
8      ; TITLE OF INVENTION: Use In Diagnostic Methods And Vaccines
9      ; NUMBER OF SEQUENCES: 98
10     ; CORRESPONDENCE ADDRESS:
11     ; ADDRESSEE: MORGAN & FINNEGAN
12     ; STREET: 345 PARK AVENUE
13     ; CITY: NEW YORK
14     ; STATE: NEW YORK
15     ; COUNTRY: USA
16     ; ZIP: 10154
17     ; COMPUTER READABLE FORM:
18     ; MEDIUM TYPE: FLOPPY DISK
19     ; COMPUTER: IBM PC COMPATIBLE
20     ; OPERATING SYSTEM: PC-DOS/MS-DOS
21     ; SOFTWARE: WORDPERFECT 5.1
22     ; CURRENT APPLICATION DATA:
23     ; APPLICATION NUMBER: PCT/US93/08849A
24     ; FILING DATE: 17-SEP-1993
25     ; PRIOR APPLICATION DATA:
26     ; APPLICATION NUMBER: US07/947,263
27     ; FILING DATE: 18-SEP-1992
28     ; ATTORNEY/AGENT INFORMATION:
29     ; NAME: William S. Fellet
30     ; REGISTRATION NUMBER: 26,728
31     ; REFERENCE/DOCKET NUMBER: 2026-4032 PCT
32     ; TELECOMMUNICATION INFORMATION:
33     ; TELEPHONE: (212) 758-4800
34     ; TELEFAX: (212) 751-6849
35     ; INFORMATION FOR SEQ ID NO: 1:
36     ; SEQUENCE CHARACTERISTICS:
37     ; LENGTH: 1693 AMINO ACID RESIDUES
38     ; TYPE: AMINO ACID
39     ; STRANDEDNESS: UNKNOWN
40     ; TOPOLOGY: UNKNOWN
41     ;
42     ;
43     ;
44     ;
45     ;
46     ;
47     ;
48     ;
49     ;
50     ;
51     ;
52     ;
53     ;
54     ;
55     ;
56     ;
57     ;
58     ;
59     ;
60     ;
61     ;
62     ;
63     ;
64     ;
65     ;
66     ;
67     ;
68     ;
69     ;
70     ;
71     ;
72     ;
73     ;
74     ;
75     ;
76     ;
77     ;
78     ;
79     ;
80     ;
81     ;
82     ;
83     ;
84     ;
85     ;
86     ;
87     ;
88     ;
89     ;
90     ;
91     ;
92     ;
93     ;
94     ;
95     ;
96     ;
97     ;
98     ;
99     ;
100    ;
101    ;
102    ;
103    ;
104    ;
105    ;
106    ;
107    ;
108    ;
109    ;
110    ;
111    ;
112    ;
113    ;
114    ;
115    ;
116    ;
117    ;
118    ;
119    ;
120    ;
121    ;
122    ;
123    ;
124    ;
125    ;
126    ;
127    ;
128    ;
129    ;
130    ;
131    ;
132    ;
133    ;
134    ;
135    ;
136    ;
137    ;
138    ;
139    ;
140    ;
141    ;
142    ;
143    ;
144    ;
145    ;
146    ;
147    ;
148    ;
149    ;
150    ;
151    ;
152    ;
153    ;
154    ;
155    ;
156    ;
157    ;
158    ;
159    ;
160    ;
161    ;
162    ;
163    ;
164    ;
165    ;
166    ;
167    ;
168    ;
169    ;
170    ;
171    ;
172    ;
173    ;
174    ;
175    ;
176    ;
177    ;
178    ;
179    ;
180    ;
181    ;
182    ;
183    ;
184    ;
185    ;
186    ;
187    ;
188    ;
189    ;
190    ;
191    ;
192    ;
193    ;
194    ;
195    ;
196    ;
197    ;
198    ;
199    ;
200    ;
201    ;
202    ;
203    ;
204    ;
205    ;
206    ;
207    ;
208    ;
209    ;
210    ;
211    ;
212    ;
213    ;
214    ;
215    ;
216    ;
217    ;
218    ;
219    ;
220    ;
221    ;
222    ;
223    ;
224    ;
225    ;
226    ;
227    ;
228    ;
229    ;
230    ;
231    ;
232    ;
233    ;
234    ;
235    ;
236    ;
237    ;
238    ;
239    ;
240    ;
241    ;
242    ;
243    ;
244    ;
245    ;
246    ;
247    ;
248    ;
249    ;
250    ;
251    ;
252    ;
253    ;
254    ;
255    ;
256    ;
257    ;
258    ;
259    ;
260    ;
261    ;
262    ;
263    ;
264    ;
265    ;
266    ;
267    ;
268    ;
269    ;
270    ;
271    ;
272    ;
273    ;
274    ;
275    ;
276    ;
277    ;
278    ;
279    ;
280    ;
281    ;
282    ;
283    ;
284    ;
285    ;
286    ;
287    ;
288    ;
289    ;
290    ;
291    ;
292    ;
293    ;
294    ;
295    ;
296    ;
297    ;
298    ;
299    ;
300    ;
301    ;
302    ;
303    ;
304    ;
305    ;
306    ;
307    ;
308    ;
309    ;
310    ;
311    ;
312    ;
313    ;
314    ;
315    ;
316    ;
317    ;
318    ;
319    ;
320    ;
321    ;
322    ;
323    ;
324    ;
325    ;
326    ;
327    ;
328    ;
329    ;
330    ;
331    ;
332    ;
333    ;
334    ;
335    ;
336    ;
337    ;
338    ;
339    ;
340    ;
341    ;
342    ;
343    ;
344    ;
345    ;
346    ;
347    ;
348    ;
349    ;
350    ;
351    ;
352    ;
353    ;
354    ;
355    ;
356    ;
357    ;
358    ;
359    ;
360    ;
361    ;
362    ;
363    ;
364    ;
365    ;
366    ;
367    ;
368    ;
369    ;
370    ;
371    ;
372    ;
373    ;
374    ;
375    ;
376    ;
377    ;
378    ;
379    ;
380    ;
381    ;
382    ;
383    ;
384    ;
385    ;
386    ;
387    ;
388    ;
389    ;
390    ;
391    ;
392    ;
393    ;
394    ;
395    ;
396    ;
397    ;
398    ;
399    ;
400    ;
401    ;
402    ;
403    ;
404    ;
405    ;
406    ;
407    ;
408    ;
409    ;
410    ;
411    ;
412    ;
413    ;
414    ;
415    ;
416    ;
417    ;
418    ;
419    ;
420    ;
421    ;
422    ;
423    ;
424    ;
425    ;
426    ;
427    ;
428    ;
429    ;
430    ;
431    ;
432    ;
433    ;
434    ;
435    ;
436    ;
437    ;
438    ;
439    ;
440    ;
441    ;
442    ;
443    ;
444    ;
445    ;
446    ;
447    ;
448    ;
449    ;
450    ;
451    ;
452    ;
453    ;
454    ;
455    ;
456    ;
457    ;
458    ;
459    ;
460    ;
461    ;
462    ;
463    ;
464    ;
465    ;
466    ;
467    ;
468    ;
469    ;
470    ;
471    ;
472    ;
473    ;
474    ;
475    ;
476    ;
477    ;
478    ;
479    ;
480    ;
481    ;
482    ;
483    ;
484    ;
485    ;
486    ;
487    ;
488    ;
489    ;
490    ;
491    ;
492    ;
493    ;
494    ;
495    ;
496    ;
497    ;
498    ;
499    ;
500    ;
501    ;
502    ;
503    ;
504    ;
505    ;
506    ;
507    ;
508    ;
509    ;
510    ;
511    ;
512    ;
513    ;
514    ;
515    ;
516    ;
517    ;
518    ;
519    ;
520    ;
521    ;
522    ;
523    ;
524    ;
525    ;
526    ;
527    ;
528    ;
529    ;
530    ;
531    ;
532    ;
533    ;
534    ;
535    ;
536    ;
537    ;
538    ;
539    ;
540    ;
541    ;
542    ;
543    ;
544    ;
545    ;
546    ;
547    ;
548    ;
549    ;
550    ;
551    ;
552    ;
553    ;
554    ;
555    ;
556    ;
557    ;
558    ;
559    ;
560    ;
561    ;
562    ;
563    ;
564    ;
565    ;
566    ;
567    ;
568    ;
569    ;
570    ;
571    ;
572    ;
573    ;
574    ;
575    ;
576    ;
577    ;
578    ;
579    ;
580    ;
581    ;
582    ;
583    ;
584    ;
585    ;
586    ;
587    ;
588    ;
589    ;
590    ;
591    ;
592    ;
593    ;
594    ;
595    ;
596    ;
597    ;
598    ;
599    ;
600    ;
601    ;
602    ;
603    ;
604    ;
605    ;
606    ;
607    ;
608    ;
609    ;
610    ;
611    ;
612    ;
613    ;
614    ;
615    ;
616    ;
617    ;
618    ;
619    ;
620    ;
621    ;
622    ;
623    ;
624    ;
625    ;
626    ;
627    ;
628    ;
629    ;
630    ;
631    ;
632    ;
633    ;
634    ;
635    ;
636    ;
637    ;
63
```

PCT-US93-08849A-1

Query Match 82.5%; Score 7439; DB 5; Length 1693;
 Best Local Similarity 81.6%; Pred. No. 0.
 Matches 197; Conservative 111; Mismatches 162; Indels 42; Gaps 7;

```

QY 1 PCTTAEQAAALAAANASALANAVVAPFLSRVOTELLINLMOPROLVPEPEVLMNPIOR 60
DB 10 PCTTAEQAAALAAANASALANAVVAPFLSRVOTELLINLMOPROLVPEPEVLMNPIOR 69
QY 61 VHNLEBOYCRABAGRCLEVEGAHPSINDPNVLRHCELRPGRDVORNYSAFTKGPAA 120
DB 70 VHNLELETCRABAGRCLEVEGAHPSINDPNVLRHCELRPGRDVORNYSAFTKGPAA 129
QY 121 CRRSALRGPRADRYCEFGFSGCNPAETGIALYSLHDMSPDVAEAMFRHGMRLYA 180
DB 130 CRRSALRGPRADRYCEFGFSGCNPAETGIALYSLHDMSPDVAEAMFRHGMRLYA 189
QY 181 LHLPEVLLPCTYHTTSTYLLIHDGRAVYVEGDTSAQYNHDSVILRAMIRTKIVGDH 240
DB 190 LHLPEVLLPCTYHTTSTYLLIHDGRAVYVEGDTSAQYNHDSVILRAMIRTKIVGDH 249
QY 241 PLVIERVRAIGHFVLLTLTADEPSPMPYVPRSTEVYRSIFGPGSPLEPSSACSTK 300
DB 250 PLVIERVRAIGHFVLLTLTADEPSPMPYVPRSTEVYRSIFGPGSPLEPSSACSTK 309
QY 301 STEHNAVPIIMRMLFGATLDDQAFCCSRMLTYLFGISYKYTVGLVANEHGNASEDL 360
DB 310 STEHNAVPIIMRMLFGATLDDQAFCCSRMLTYLFGISYKYTVGLVANEHGNASEDL 369
QY 361 TAVITRAAYLTICHORLYRQAISKGMRLGVEHAQKFTIRLYSMLEKSGRDYIPROLO 420
DB 370 TAVITRAAYLTICHORLYRQAISKGMRLGVEHAQKFTIRLYSMLEKSGRDYIPROLO 429
QY 421 EYAQCRMLWSAGPHLDPRVLVFEDESVCRCRTFLKAVAKFCCFMKLOQECTCELEPAE 480
DB 430 EYAQCRMLWSAGPHLDPRVLVFEDESVCRCRTFLKAVAKFCCFMKLOQECTCELEPAE 489
QY 481 GLVGDHGHNEAVEGSEVDPAPBAHLVDSGTAVHGHOLEALYRALNVODILAAARSRLT 540
DB 490 GLVGDHGHNEAVEGSEVDPAPBAHLVDSGTAVHGHOLEALYRALNVODILAAARSRLT 549
QY 541 ATVELEVASPDLRECRVTLNKTFRTTVVDGHLLEANGPEQOYVLSFDSQSMGASHSLT 600
DB 550 ATVELEVASPDLRECRVTLNKTFRTTVVDGHLLEANGPEQOYVLSFDSQSMGASHSLT 609
QY 601 YELTPAGLOVRISSNGLDCTATFPPEGADSAAPGEVAACFCSALYRYNRFTORHSLTGLM 660
DB 610 YELTPAGLOVRISSNGLDCTATFPPEGADSAAPGEVAACFCSALYRYNRFTORHSLTGLM 669
QY 661 LHPGGLIGFPFPGHIMESANPECGEGTLTYRTWS-----TSGFSSDSPPE 709
DB 670 LHPGGLIGFPFPGHIMESANPECGEGTLTYRTWS-----TSGFSSDSPPE 724
QY 710 AAAPMAATPGLPHSTPPVSDIWLVLPPSEEFQVDAAPV-PPAPDPA-GLPGFVVLTPPP 767
DB 725 PSIPSAAPR-----TP-----AAPIPPAPDPSPLTSAAPAREPAP 761
QY 768 PPPVHKPSLP-PPSRNRRLITYYPDGAKYAGSLFSDCDMLVNASNPGHREGGCLCHAF 826
DB 762 GATAPAPATHTOTARHRLRLLFTYPPDOSKVFAGSLFESTCTWLNVASNVHRGGGCLCHAF 821
QY 827 YORPEAFYPTFEIIMEGLAAATLPPRPIIHAVAPDYRVEONPKRIEALYREFCSRGRA 886
DB 822 YORPEAFYPTFEIIMEGLAAATLPPRPIIHAVAPDYRVEONPKRIEALYREFCSRGRA 881
QY 887 AYPPLGSGIYQVVSLSFDAMERNHRPGDELXLTPEPAMFEANKPAQVLLTTEDTANT 946
DB 882 AYPPLGSGIYQVVSLSFDAMERNHRPGDELXLTPEPAMFEANKPAQVLLTTEDTANT 941
QY 947 ANLALEIDATVGRACAGCTISPGIVHQFTAGVPGSGKSRSIQOQDVVVVPTREIR 1006
DB 942 ANLALEIDATVGRACAGCTISPGIVHQFTAGVPGSGKSRSIQOQDVVVVPTREIR 1001

```

```

DB 942 ANLALEIDATVGRACAGCTISPGIVHQFTAGVPGSGKSRSIQOQDVVVVPTREIR 1001
QY 1007 NSWRRRGFAAFTPHTAARTYIGRRVYIDEAPSLPPLLHLMORASSVHLGDPNOIPAI 1066
DB 1002 NAMRRRGFAAFTPHTAARTYIGRRVYIDEAPSLPPLLHLMORASSVHLGDPNOIPAI 1061
QY 1067 DEHAGLVPAIRBELAPTSMMVYTHRCPADVCELLRGAYPKIQTTSRVLSLFWNEPAIG 1126
DB 1062 DEHAGLVPAIRBELAPTSMMVYTHRCPADVCELLRGAYPKIQTTSRVLSLFWNEPAIG 1121
QY 1127 OKLVYQAANKANPGALITYHEAOGAFTEFTTTIATADAGLIQSSRAHAIVALTRTEKC 1186
DB 1122 OKLVYQAANKANPGALITYHEAOGAFTEFTTTIATADAGLIQSSRAHAIVALTRTEKC 1181
QY 1187 VILIDAGLREVGISDVYINNFFLAGCEYGHXHPSVIYPPGNPNQCTLOAPSPCOISA 1246
DB 1182 VILIDAGLREVGISDVYINNFFLAGCEYGHXHPSVIYPPGNPNQCTLOAPSPCOISA 1241
QY 1247 YHQLAEELGHRAPVAAVLPCEPELBGLLYMPOELTVSDSVLVFELTDIVHCRMAAPSO 1306
DB 1242 YHQLAEELGHRAPVAAVLPCEPELBGLLYMPOELTVSDSVLVFELTDIVHCRMAAPSO 1301
QY 1307 KRAVUSTLVGRTRTKLYLEAHSVDRESLARIPTIGVQATCTCLYELVEAMVEKGD 1366
DB 1302 KRAVUSTLVGRTRTKLYLEAHSVDRESLARIPTIGVQATCTCLYELVEAMVEKGD 1361
QY 1367 GSAVLELDLCNDRVSRITFEFOKXCNKFTTGETIAGHKVCGGISAMSKTCALPGMPRAI 1426
DB 1362 GSAVLELDLCNDRVSRITFEFOKXCNKFTTGETIAGHKVCGGISAMSKTCALPGMPRAI 1421
QY 1427 EKELLALPPIETGYDAVEESVFAAASVAGSCMVEENDESEEDSTONNESLGEVME 1486
DB 1422 EKELLALPPIETGYDAVEESVFAAASVAGSCMVEENDESEEDSTONNESLGEVME 1481
QY 1487 ECGMPQMLIRLYHLVASNTILOAPKESLKGPMKHSGECEGLTLMNVMMATIAHCYER 1546
DB 1482 ECGMPQMLIRLYHLVASNTILOAPKESLKGPMKHSGECEGLTLMNVMMATIAHCYER 1541
QY 1547 DFRVAAFKDDSVVLCSDYRSHNAALJAGGKLKLVDRYRPIGLTAGVVAAGLCTLPD 1606
DB 1542 DFRVAAFKDDSVVLCSDYRSHNAALJAGGKLKLVDRYRPIGLTAGVVAAGLCTLPD 1601
QY 1607 VYRFAGLRSEKMGWGPBEBEOLRLAVCDFLRLTNVAQVVDVVSRYGVSPGLVHNL 1666
DB 1602 VYRFAGLRSEKMGWGPBEBEOLRLAVCDFLRLTNVAQVVDVVSRYGVSPGLVHNL 1661
QY 1667 GMLQTTADGKAFHTETIKYVLDLNSIIORVE 1698
DB 1662 GMLQTTADGKAFHTETIKYVLDLNSIIORVE 1693

```

RESULT 7
 PCT-US93-08849-1
 ; Sequence 1, Application PC/TUS9308849
 ; GENERAL INFORMATION:
 ; APPLICANT: Tsarev, Sergei A., Emerson,
 ; APPLICANT: Suzanne U., Purcell, Robert H.
 ; TITLE OF INVENTION: Recombinant Proteins Of
 ; TITLE OF INVENTION: A Pakistani Strain Of Hepatitis E And Their
 ; NUMBER OF INVENTION: Use In Diagnostic Methods And Vaccines
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MORGAN & FINNEGAN
 ; STREET: 345 PARK AVENUE
 ; CITY: NEW YORK
 ; STATE: NEW YORK
 ; COUNTRY: USA
 ; ZIP: 10154
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: FLOPPY DISK
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: WORDPERFECT 5.1

US-08-478-507-2
; Sequence 2, Application US/08478507
; Patent No. 6120988
; GENERAL INFORMATION:
; APPLICANT: Reyes, Gregory R
; APPLICANT: Yarbough, Patrice O
; APPLICANT: Bradley, Daniel W
; APPLICANT: Krawczynski, Krzysztof Z
; APPLICANT: Tam, Albert
; APPLICANT: Fry, Kirk E
; TITLE OF INVENTION: DNA Sequences of Enterically Transmitted
; TITLE OF INVENTION: No. 6120988-A/No. 6120988-B Hepatitis Viral Agent
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,507
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/279,823
; FILING DATE: 25-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/681,078
; FILING DATE: 05-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/505,888
; FILING DATE: 05-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/420,921
; FILING DATE: 13-OCT-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/367,486
; FILING DATE: 16-JUN-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/336,672
; FILING DATE: 11-APR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/208,997
; FILING DATE: 17-JUN-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 324-0880
; TELEFAX: (650) 324-0960
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 431 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-478-507-2

Query Match 22.8%; Score 2057; DB 3; Length 431;
Best Local Similarity 88.4%; Pred. No. 1.8e-172;
Matches 381; Conservative 26; Mismatches 24; Indels 0; Gaps 0;

QY 1257 RAAPAAVILPCPELEGGILMPDELTVSDSLVLELTIIVICRMAAPSGRAVAVSTLVG 1316
DB 1 RVPVAAVILPCPELEGGILMPDELTVSDSLVLELTIIVICRMAAPSGRAVAVSTLVG 60

QY 1317 RYGRRTKLYEAASDVRESLARIPTIGPQATTCCLAYELVEAMVEKGDSAVLELDLC 1376
DB 61 RYGRRTKLYEAASDVRESLARIPTIGPQATTCCLAYELVEAMVEKGDSAVLELDLC 120
QY 1377 NRDSRIITFPQKXCNKFTTGETIAGKVGQGISAMSKTICALFGPMFRAIEKLTALLPP 1436
DB 121 NRDSRIITFPQKXCNKFTTGETIAGKVGQGISAMSKTICALFGPMFRAIEKLTALLPP 180
QY 1437 NIFGDAVEESVPAAVASGCMVFENDESEPDSTONNFSIGLECVNMEEGMPWLIR 1496
DB 181 GVFYGAEPDVFSAVAALAKASVWFENDESEPDSTONNFSIGLECVNMEEGMPWLIR 240
QY 1497 LYHLVRSAMTILAPKESLGFKKHSGEPGLTLMNTVMMAALIAHCEFRDPRVAAFKGD 1556
DB 241 LYHLVRSAMTILAPKESLGFKKHSGEPGLTLMNTVMMAALIAHCEFRDPRVAAFKGD 300
QY 1557 DSVILCSIDYRQSRMAALIAAGGLKIKVDYRPIGLYAGVVAAGLGLDPDVYRFAGRISE 1616
DB 301 DSVILCSIDYRQSRMAALIAAGGLKIKVDYRPIGLYAGVVAAGLGLDPDVYRFAGRISE 360
QY 1617 KMWGPPEAEQDLRLAVCPFLGLTVNAOVYCVDVYSRVYGVSPGLVHNLIGMLQTIADK 1676
DB 361 KMWGPPEAEQDLRLAVCPFLGLTVNAOVYCVDVYSRVYGVSPGLVHNLIGMLQTIADK 420
QY 1677 AHFTETIKPVL 1687
DB 421 AHFTESVKPVL 431

RESULT 9
US-09-128-275A-2
; Sequence 2, Application US/09128275A
; Patent No. 6229005
; GENERAL INFORMATION:
; APPLICANT: Reyes, Gregory R
; APPLICANT: Yarbough, Patrice O
; APPLICANT: Bradley, Daniel W
; APPLICANT: Krawczynski, Krzysztof Z
; APPLICANT: Tam, Albert
; APPLICANT: Fry, Kirk E
; TITLE OF INVENTION: DNA Sequences of Enterically Transmitted
; TITLE OF INVENTION: No. 6229005-A/No. 6229005-B Hepatitis Viral Agent
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/128,275A
; FILING DATE: 03-AUG-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/279,823
; FILING DATE: 25-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/681,078
; FILING DATE: 05-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/505,888
; FILING DATE: 05-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/420,921
; FILING DATE: 13-OCT-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/367,486
; FILING DATE: 16-JUN-1989

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/336,672
; FILING DATE: 11-APR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/208,997
; FILING DATE: 17-JUN-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Petithory, Joanne R.
; REGISTRATION NUMBER: 42,995
; REFERENCE/DOCKET NUMBER: 4600-0183.24
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 324-0980
; TELEFAX: (650) 324-0960
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 431 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-128-275A-2

Query Match      22.8%; Score 2057; DB 4; Length 431;
Best Local Similarity 88.4%; Pred. No. 1.8e-172;
Matches 381; Conservative 26; Mismatches 24; Indels 0; Gaps 0;

QY 1257 RPAVAAVLPCEPLEGGLIMPELVSDSVLVEFLTDIYHGMARPSQKAVLSTLVG 1316
      |||||||
      1 RPAVAAVLPCEPLEGGLIMPELVSDSVLVEFLTDIYHGMARPSQKAVLSTLVG 60

QY 1317 RYGRRTLYEAHSDVRESLARFPTIGPVQATCELYELVEAVEKGGDSAVLELDLC 1376
      |||||||
      61 RYGRRTLYEAHSDVRESLARFPTIGPVQATCELYELVEAVEKGGDSAVLELDLC 120

QY 1377 NRDVSRITTFQKXCNKFTTETIAGKVGOGISAMSKTFCLFGPWRAIEKETALLLP 1436
      |||||||
      121 NRDVSRITTFQKXCNKFTTETIAGKVGOGISAMSKTFCLFGPWRAIEKETALLLP 180

QY 1437 NIFPGDAVESVRAAVSAGSGCMVFENDSEPDSTONNSLGLCYVMECCGMPWLIR 1496
      |||||||
      181 NIFPGDAVESVRAAVSAGSGCMVFENDSEPDSTONNSLGLCYVMECCGMPWLIR 240

QY 1497 LYHLVRSAMTLOAKESLKGKWKHSGEPGTLNNTWNNMAIIHGYEERFVRFAAFKGD 1556
      |||||||
      241 LYHLVRSAMTLOAKESLKGKWKHSGEPGTLNNTWNNMAIIHGYEERFVRFAAFKGD 300

QY 1557 DSVVLCSDYRSRMAALLIAGCGILKLVYRPIGLYAGVVAAPGLTLPDVVRFAGRLSE 1616
      |||||||
      301 DSVVLCSDYRSRMAALLIAGCGILKLVYRPIGLYAGVVAAPGLTLPDVVRFAGRLSE 360

QY 1617 KMGCPGERAEQLRLAVCDLRLGLITNVAQCVDVVSRYVSGPLVHNLIGMLQTIADGK 1676
      |||||||
      361 KMGCPGERAEQLRLAVCDLRLGLITNVAQCVDVVSRYVSGPLVHNLIGMLQTIADGK 420

QY 1677 AHFTETIKPVY 1687
      |||||||
      421 AHFTETIKPVY 431

RESULT 10
US-08-485-355B-40
; Sequence 40, Application US/08485355B
; Patent No. 6177075
; GENERAL INFORMATION:
; APPLICANT: Christian, P. D., Gordon, K. H.J., Hanzlik, T. N.
; TITLE OF INVENTION: Insect Viruses and Their Uses in
; Protecting Plants
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr Hobbach Test Albritton & Herbert LLP
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California

```

```

; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,355B
; FILING DATE: 07-Jun-1995
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/440,522
; FILING DATE: 12-MAY-1995
; APPLICATION NUMBER: US 08/089,372
; FILING DATE: 08-JUL-1993
; APPLICATION NUMBER: AU PL4081/92
; FILING DATE: 14-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-58631-2/RFT/DSS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1704 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 40:
; US-08-485-355B-40

Query Match      5.7%; Score 515; DB 4; Length 1704;
Best Local Similarity 21.7%; Pred. No. 1e-35;
Matches 397; Conservative 179; Mismatches 553; Indels 700; Gaps 84;

QY 13 AAASALANAVVVR-----PFLSVQETELINLMOPROLVFRPEVL--WNHPI---Q 59
      |||
      13 AAASALANAVVVR-----PFLSVQETELINLMOPROLVFRPEVL--WNHPI---Q 68

QY 60 RVINELEQYCRARAGCLEYGAHPRS--INDPNV---LHRCFLRPVGRDYOVRKXA 112
      |||
      69 RVINELEQYCRARAGCLEYGAHPRS--INDPNV---LHRCFLRPVGRDYOVRKXA 126

QY 113 PTRGPANCRSALRGLP-----ADRTYCPDGSRCACFAAETGVALXSLHDL 160
      |||
      127 LE-----SRSVATGREPFRKADSLANGIASRTPCYDVGSCAFKSRVGIANSLTDV 179

QY 161 WPAVDAEAMARHGXTRLYAALHLPPEVLLPGYHTTYSYLLIH-----DGD 206
      |||
      180 TLEELANFENHGLHMAFAHMBELLMYNNVNAELGYFRHYIEEPMMAVKDCAFQDGD 239

QY 207 -----RAVVTYEGDPSAG--YNHDVSLIRAMIRTK 235
      |||
      240 LRLHPELDELINESQERRIERILARGSYSRRAVIFSGDDWDGDAYLHDFHTMLAYL---- 295

QY 236 IVGDHP-----LYIERRALGCHFVLLTLAAPESPYPYPRSTYVYRSIFGPGGS 289
      |||
      296 LVKNVPTPEGFSLHIEVORHGSSIELKTRAP-----PGDMALAVNVPRTSGLCRI 347

QY 290 PSLE--PSACSTKSTFHAVPHIMDRMLFEATLDDQAF--CSRLMTYLRGISYKVTYGA 346
      |||
      348 PNIFYYADASTGTEKTLTISQHKVNMMLNFQGTREKELVDMTYLMSRARLRALIVYAS 407

QY 347 LVANEGWNASDALTAXITAAVLTICHQRYLRTQAIKSMRRLGEVHAQKFTIRLYSWLF 406
      |||
      408 EVTESWNIISPADLVRTVVSILYHITER-----BRAAVA----- 442

QY 407 EKSGRDYIPGQLOFYACQKRWLSAGFHLDRVLFVDESVCRCRRTFLKKYAGKFCFCMR 466

```


Matches 454; Conservative 201; Mismatches 714; Indels 995; Gaps 104;

```
QY 3 ITTAIEQALAAANSAANVAVVRFSLRYQTEILINLMQPROLVFPEVLMNHPIORVI 62
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 40 VVTAACKRAIV-----AVIPRVFTQMVS-----DHP--AL 69

QY 63 HNELEQYCARAGRCLEVGANPRS-----INDPNVLHRCFLRPVGHVORWYS----- 111
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 70 H-AISYTRR--HWIEMG--PKFALHVLIDPSGL-----LREVAVERRWALCHIRT 118

QY 112 -----APTRGAAN-----CRSALRGLP-----ADRTYC-FDGF 141
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 119 ARKLATALFETASEAMHADVC--ALRGAPSGPFVHPEDVPHGKRAVADRCLLYTPM 175

QY 142 SRCAFAGFVALYSLHDLMPADVA-----E 167
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 176 QKELMRTIDATLLVAVDIMPVALAAHVGDWDGLIAMHLHDGGCPADCGAGAGPTP 235

QY 168 AMARHGXTRLYAALHLPPEVLLPRTYHTSYLLIHGDRAVVTYEGDTSAGYNHDVSL 227
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 236 GYTRPCTRIYQVL-----PDTAHGRLXKCGPRMLTRDCANAEISWEVAQHCQD----- 286

QY 228 RANIRTKIYGDHPVIERVRAIGCFVLLLTAAPESPMPVYVPRSTEVYYSIFPG 287
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 287 -----ARVAVRCTLPJRHVSLSQS-----ARVR----- 311

QY 288 GSPSLPFSACSTSTHNAVVHIMDRMLFGLALDQAFCCSLMNTYLRISIKVYVYAL 347
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 312 -LPDLVHLAEVGRWFSLSRPVFORMLSYCKTLPDAVYSEVYFKRNALCHSITLAGN 370

QY 348 VANEGN-----ASEDALTAATAVYLTICHOYRLTQAIKSGMRILGEHAOKFITRLYS- 403
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 371 VLOEGKRGTCABEDALCA-----YAFRAWGSNALAGIMKAKCAADSLV 417

QY 404 --WLFKSGHDYIPGROLOFYAOCRMLSA-----GFHDPRLVFEDESVCRCRTF 453
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 418 AGWL-----DTI-----WDAIKRFLGSVPLAERMEWEOADAAVAHPD----- 454

QY 454 LKVVAKFCFCFMWLGEOECTFLPAPGLVGDGHNEAVEGSEVDPAPAEHLIDVSTYA 513
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 455 -----RGPLEDGGRLDTVQ-----PKSPREPEIATWT 484

QY 514 VHGHOLE--ALYRALNVPOD-----IARAS 537
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 485 VHAASEDRHCACAPRCDDVPRERSAPAGOPDDALLPFWLFARRLRCHEMFELARAR 544

QY 538 RLTAIVELVASPDLECRVYL-----GNKFRFTYVDGANLEANGPEOY 581
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 545 ADTAAAPAPRAPAPRAPYPTVLYRHPAHNGPWLTLDEGEADALVLCDDPLGQPLRGPERH 604

QY 582 VLSFDSRQSMGAGSHSLYELTRPAGLO--VRISNSGLDCTATFFPREGAPSAAPGEVAAF 639
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 605 -----FAAGAHMCA--QARGLOAEVRV-----VPPEREMAGGG--ARA 639

QY 640 CSALYYNRFNFTORHSLTGLIMLPEGLTG--IFPPFSGHIMESANPFCGGTLYTTR 695
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 640 MKKFFFGCAMAQR-----LIGBRAVAMHLPYTDGDPQLI-----ALAKRT 679

QY 696 WTSIGSSSDPS--PEEAPAPMAAT-----PGLP-----HS 724
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 680 LAQOGAALALSVYDLPGGAFDANAVTAARVAGPRQSAASAPPPGDPRRARRSQRHS 739

QY 725 -----PPVSDIWLPPSEFOVDAPVPPAPRAGLPQGVLTLPVPPVHK----- 773
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 740 DARGTTPPAPADPPP-----APSPAPRADDPV-----PIPAGADARDAELE 787

QY 774 -----PSIPPSRNR-----RLTYYPDGAVYVYASLFEPCDWL 808
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 788 VACEPSGPTSTRADPDSDIVESYARAGPVHLRYVDIMDPGCKV-----Y 835

QY 809 VNASNNGHPRGGGLCHAFYQRPPEAFYPTFEIMREGIAA--YTLTPRP----- 854
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 836 VNAANEGLLAGSGVCAIFANATAA-----LAANCRLAPCPTGEAVATPGHGC 884
```

```
QY 855 -----IIHAVP-----DYVEONPKRLAEAYRE-----TCSRGETAYPLLSGIT--QV 898
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 885 GYTHIIHAIVAPRRRDRPALEBGBALLERAYRSTIVALAARRMACVACPILGAGVYWSA 944

QY 899 PVSLSPDAMERNRHPGDELY-----TEPANMFE 928
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 945 AESIRALAAATRTPEPVRSVSLCHCPDRATLLHASVLYGAGLAARRVSPPTETLAS-CP 1003

QY 929 ANKPAQVLTITTEDTANTANLALEIDATEVGRACACTIS----- 969
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1004 AGDPRGPAOSASBPATPLG-----DATAPEPRGCGCELCRYTRYTNDRAVYVLMERD 1058

QY 970 -----PGIV-----HQF----- 977
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1059 RGATSWAMRLPEVYVYVGREHLATHFPLNHSVYLKPAEVRPPRGMGSDMMRCRWGMQP 1118

QY 978 -----TAGVSGSGKSRSIQGDVDVVVYVTRELRN--SWRRRGFAFTPTAA 1023
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1119 VRCTPSMAHALCRTGVPPRASTRG--GELDPNTCMLRAANVQAARACGAYTSAGCP 1175

QY 1024 RVTIGR-----RVVIDEAP 1037
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1176 KCAVGRALSEARTHEDEAFALSQKMSASHADSPDGTGDPDPLMETVGCACSRVWGESEH 1235

QY 1038 SUPPHLLHMOA-----SSVHL-----GDPNQI--PAI----- 1066
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1236 EAPPDHLVSLHRAPNPGMVLEVARPREGNPTGHEVCAVGGPRVSDRPHLMVAVP 1295

QY 1067 -----DFE-----HAGLVAIPELAPTS--WMXYVT-- 1090
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1296 LSRGGCATDEGLAQAYYDLEVRRLGDANARALASVQPRKGPYINIRVNMMAAGA 1355

QY 1091 -----HRCPA-----DYCELIR-----GAPKI-- 1108
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1356 GKTRIIAAFTREDLYCPTNALLHETQAKIRARDIDIKNAATYERRLTKPLAAYRITYI 1415

QY 1109 -----QTSRYL-----RSLE-WNEP- 1123
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1416 DEAFITGEGEYCAVVASQTTAEVLCVGDRCQGPYANNCRTPYVDPMPRTSRHTWEPD 1475

QY 1124 -----AIGQKLYTQAKAANGALTVHE 1147
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1476 CMAARLRAGLDYDIEGRRTGFACNLMDRGQVLDHLAFSEHY--RLHAGIRATYVRE 1533

QY 1148 AOGATFETTTIATDARGL--IQSSRAHAIVALTRETEKCVILD-APGLLEVGISDIYV 1205
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1534 AOGMSVGTACIHVGRDGTVALATRLDALVSLTRASDALYHLELEDGSLRAAGLS---- 1589

QY 1206 NNFFLAGGEV-----GNHR-----PSVIPRGPNQNLGLOAFPPSCQISATVHOL 1250
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1590 --AFLDGALAELEKEVPAGIDRVAAVEQAPPLP--PADGIEPADVDPFPCRTLEELV 1644

QY 1251 ABEILGHRPAPVAVLPCCPELEQGLYMPQELTVSDSVYFELTDIYHCMAAPSOGRKAV 1310
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1645 FGKAGH--PHYADLNKRVYTGGEREVKM--RISRLILKNNTHEM-----PGTR-V 1689

QY 1311 LSTLVGRYGRRTKLYEAAH--SDVRSLARFIP--TIGPVQATY-----CELYEL 1356
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1690 LSAVCA-----VRRYRAGEDGSLRTAVARQHPRPFPQIIPRPVYTAGVAQEWMTYLRER 1744

QY 1357 VEAMVEKGOGSVALEL-----DLCNRDVSRTTFQKCKNKF-----T 1394
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1745 IDLTDVYTMGVARELTDYRARRYPEIFAGMCTAOSLSVPAPLAKATLKCVDAALGRDPT 1804

QY 1395 TGETIAGKVGQGISAMSKFPCALFGPWFRAIKETIATLPINIFYGDAVEESVFAAAYS 1454
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1805 EDCHAQKAGKGLIIRMAKMWQVMSHPRAIQIIMRALRPOPIVAAGITTEBVDAWMQ 1864

QY 1455 GAGSCWFFENDESEPDSTONNES-----LGEVYMEECGNPQWLRILYHLVRS 1503
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1865 AHYTTNAIEVDFTFEFDNMQTLATRDVELTISAALLGHPCA--BD-----YRALRA 1912
```

Accession	Sequence	Position
QY 1504	-AMTLQAKREKCKEFMKHSGEGGTLIMNTVMNMAIIARCEER----	DERIVA-FKGD 1557
Db 1913	GSCTLRBELGSTLEGCERTSGEPATLIHNT-----VAMCAMRMVPRKGVNAGCTFGSD	1967
QY 1558	SVVLCSDDROSRNAAL-----IAGGCLKLVDYRIGIYAGVYVAPGIGLTPDVR	1609
Db 1968	WVIFLPRGANS--ALKWPAEVLGEFGRIPIVKHSTPTSPFGC-HGTNAGLEHDVMH	2023
QY 1610	FAGRLSEKNMGPGPERAEQILRLAVCDPLRGTLINVAOCYDVSVRYGVSPGLVHNLIGML	1669
Db 2024	QAIIVLCRRF--PBDVLEEQVALDLRLRLR-----YVALP-----	2057
QY 1670	QTIADGKAHFETIKPILDTNSI	1693
Db 2058	DTVAANAAYIDYSAEKVLAIIVREL	2081

RESULT 12
 US-07-876-941A-31
 Sequence 31, Application US/07876941A
 Patent No. 5885768
 GENERAL INFORMATION:
 APPLICANT: Reyes, Gregory R.
 APPLICANT: Bradley, Daniel W.
 APPLICANT: Tam, Albert W.
 APPLICANT: Mitchell, Carl
 TITLE OF INVENTION: Hepatitis E Virus Peptide Antigen and
 TITLE OF INVENTION: Antibodies
 NUMBER OF SEQUENCES: 76
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Dehlinger & Associates
 STREET: 350 Cambridge Avenue, Suite 250
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94306
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/876, 941A
 FILING DATE: 01-MAY-1992
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 822, 335
 FILING DATE: 17-JAN-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 505, 888
 FILING DATE: 05-APRIL-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 420, 921
 FILING DATE: 13-OCTOBER-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 367, 486
 FILING DATE: 16-JUNE-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 336, 672
 FILING DATE: 11-APRIL-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 208, 997
 FILING DATE: 17-JUNE-1988
 ATTORNEY/AGENT INFORMATION:
 NAME: Sholtz, Charles K.
 REGISTRATION NUMBER: 38, 615
 REFERENCE/DOCKET NUMBER: 4600-0093.33
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 324-0880
 TELEFAX: (415) 324-0960
 INFORMATION FOR SEQ ID NO: 31:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 78 amino acids

```

?      TYPE: amino acid
?      STRANDEDNESS: single
?      TOPOLOGY: unknown
?      MOLECULE TYPE: Protein
?      HYDROTHERMAL: NO
?      ANTI-SENSE: NO
?      ORIGINAL SOURCE:
?      INDIVIDUAL ISOLATE: Fig. 11, ORF 1, aa 1285-1362Z
?      OS-07-876-941A-31

```

Query Match	4.0%;	Score 365;	DB 2;	Length 78;
Best Local Similarity	92.2%;	Pred. No. 9.2e-25;		
Matches 71; Conservative	2;	Mismatches 4;	Indels 0;	Gaps 0;

[illegible]

```

01      13
02      US-09-194-613-5
03      ; Sequence 5, Application US/09194613
04      ; Patent No. 6251654
05      ; GENERAL INFORMATION:
06      ; APPLICANT: GORDON, Karl H.
07      ; APPLICANT: HANZLIK, Terry N.
08      ; TITLE OF INVENTION: MODIFIED SMALL RNA VIRUSES
09      ; NUMBER OF SEQUENCES: 22
10      ; CORRESPONDENCE ADDRESSES:
11      ; ADDRESSEE: McDermost, Will & Emery
12      ; STREET: 600 13th Street, N.W.
13      ; CITY: Washington
14      ; STATE: D.C.
15      ; COUNTRY: USA
16      ; ZIP: 20005-3096
17      ; COMPUTER READABLE FORM:
18      ; MEDIUM TYPE: Floppy disk
19      ; COMPUTER: IBM PC compatible
20      ; OPERATING SYSTEM: PC-DOS/MS-DOS
21      ; SOFTWARE: Patent In Release #1.0, Version #1.30
22      ; CURRENT APPLICATION DATA:
23      ; APPLICATION NUMBER: US/09/194,613
24      ; FILING DATE: 30-NOV-1998
25      ; CLASSIFICATION: 424
26      ; ATTORNEY/AGENT INFORMATION:
27      ; NAME: Bucca Ph.D., Daniel
28      ; REGISTRATION NUMBER: 42,368
29      ; REFERENCE/DOCKET NUMBER: 50179-061
30      ; TELECOMMUNICATION INFORMATION:
31      ; TELEPHONE: 202-756-8000
32      ; TELEFAX: 202-756-8087
33      ; INFORMATION FOR SEQ ID NO: 5:
34      ; SEQUENCE CHARACTERISTICS:
35      ; LENGTH: 1233 amino acids
36      ; TYPE: amino acid
37      ; STRANDEDNESS: single
38      ; TOPOLOGY: Linear
39      ; MOLECULE TYPE: protein
40      ; OS-09-194-613-5

```

Query Match	4.0%;	Score 360;	DB 4;	Length 1233;
Best Local Similarity	24.6%;	Pred. No. 2.6e-22;		
Matches 223;	Conservative 106;	Mismatches 356;	Indels 220;	Gaps 44;

QY 830 FPEAFYPTEFIMREGLAAYLTTPRIITHAVADYRVEQNPKRLAAYRETCSRGTAAYP 889
| : | : : : | : | : | : | :
Db 430 FQDKFQIDEEVVRKRYLE-CLKAQPPIHA---DKVNCETKR----YNPTVAEFGPNKHL 481

QY 1345 -----PVOATTCELYELVAME--KGODGSAVLE-----LDLCNRD 1379
Db 2116 CFRKYACNDEYMEFAKPIRITTT-----EFVTAVARLKGPRAALFAKTHNLVPLQEP 2171
QY 1380 VSRITFFQKXCKNFTTG--ETIAHGKVGOCISMSKTFCALFGPWERAIEKEITALLPPN 1437
Db 2172 MDRFVDMKRDVAVTPGTKHTERPKV-QVIAAEPLATAYLCGIHRELVRRLTAVLLPN 2230
QY 1438 I---FYGDAYEESVFAAAYSGAGSCWFEENDESEPDSTONNFSLGLECVYMECGMPWL 1494
Db 2231 IHTLFMSAEFDPAITIAEHFKQGD-PVLETDIASFDSODDAMALGLMLTEDIGVDPL 2289
QY 1495 IRL---YHLVRSAMILQAPKESLKGFWKKHSGEPGLLWNTFYWNAITAHCE--FRDF 1548
Db 2290 LDIECAFGEISSTHLP.TGTRFRFGAMK--SCMFLLFVNTVYLVNVIASRVLEERLKS 2347
QY 1549 RVAAFKGDSDV--LCSDYRQSRNAALIAGCGLKLK----VDYRPIGLYAGVVAAPGL 1601
Db 2348 KCAAFIGDNIIGHVYSDKEMAERCATWL--NMEVKIIDAVIGERPPYFCGGFILDV 2404
QY 1602 GTLPDVVRFAGRLSE--KMWGPGP---ERAEQRLAVCD-----FLRGLTNVAQVCYDV 1651
Db 2405 TS--TACRYADPLKRLFKLGLKPLPADEQDEDRRALDDETKAMFRVGIITDILAVAVATR 2462
QY 1652 SRVYGVSPGLVHNLIGMLQTIADGKAHF 1679
Db 2463 YEVDNITPVLL-----ALRTFAQSKRAF 2485

Search completed: March 7, 2002, 14:02:59
Job time: 62 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2002, 14:01:57 ; Search time 29.2 Seconds
(without alignments)
4429.602 Million cell updates/sec

Title: US-09-468-147-91

Perfect score: 9013

Sequence: 1 PGITTAEOAALAAANSALA.....FTETIKPVLDLNTSIQRIE 1698

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	7448	82.6	1693	1	MNWMHE	genome polyprotein
2	7433	82.5	1691	1	A44212	genome polyprotein
3	1368	15.2	290	2	C48547	nonstructural prot
4	763	8.5	152	2	B38196	probable RNA-direc
5	391.5	4.3	146	1	WMTMS2	186k protein - cuc
6	390.5	4.3	2115	2	S36480	nonstructural prot
7	374.5	4.2	2205	1	MNWMRN	non-structural poly
8	372	4.1	141	2	A38196	non-structural hyp
9	349	3.9	1844	2	S01956	hypothetical prote
10	346	3.8	1985	2	SI9151	hypothetical prote
11	324	3.6	1616	1	JQ2144	183k protein - tom
12	323	3.6	1615	1	WMTW8T	180k protein - tom
13	320	3.6	1597	2	S65053	genome polyprotein
14	314	3.5	1601	2	S48699	178k protein - tob
15	313	3.5	1611	1	WMTMPV	183k protein - pep
16	311	3.5	1748	1	J01555	genome polyprotein
17	310	3.4	1844	1	RRPFTM	genome polyprotein
18	305.5	3.4	1776	1	RRWPYM	genome polyprotein
19	296	3.3	1615	1	WMTM18	183.3k protein - t
20	293.5	3.3	1839	1	RRWPEM	genome polyprotein
21	293	3.3	1874	1	J00533	genome polyprotein
22	260	2.9	1608	1	WMTMGM	183k protein - tob
23	243.5	2.7	1707	2	S01865	genome polyprotein
24	243.5	2.7	1884	1	A45353	genome polyprotein
25	242.5	2.7	1885	2	JQ2183	hypothetical 216.5
26	227	2.5	1112	2	S49432	replicase 126k - o
27	226.5	2.5	1766	2	S03701	141k protein - pea
28	218	2.4	2512	1	MNWS	nonstructural poly
29	211	2.3	2514	1	MNWS82	nonstructural poly

30	207	2.3	2514	1	MNWMN2	nonstructural poly
31	202.5	2.2	2492	1	C44213	nonstructural poly
32	202	2.2	1729	2	A49282	fusion protein 1a/
33	201.5	2.2	1707	2	S77910	hypothetical prote
34	201.5	2.2	1718	1	JQ1734	genome polyprotein
35	200	2.2	2431	1	MNWSF	nonstructural poly
36	194	2.2	1968	1	PN0093	genome polyprotein
37	184.5	2.0	3175	1	RRWVEV	genome polyprotein
38	184	2.0	1707	2	S77908	hypothetical prote
39	183	2.0	1116	2	S26358	hypothetical prote
40	177.5	2.0	2492	1	MNWMYD	nonstructural poly
41	176	2.0	1385	2	S34230	156k protein - pla
42	174.5	1.9	2492	1	A44213	nonstructural poly
43	172.5	1.9	2493	2	S26372	nonstructural poly
44	169.5	1.9	2493	2	S72349	nonstructural poly
45	167	1.9	810	1	P2WMBB	2a protein - broad

ALIGNMENTS

RESULT 1

MNWMHE

genome polyprotein - hepatitis E virus (strain Burma)

N:contains: RNA-directed RNA polymerase (EC 2.7.7.48)

C:Species: hepatitis E virus

C:Date: 30-Sep-1992 #sequence-revision 30-Sep-1992 #text-change 23-Jul-1999

C:Accession: A40778; A48547

R:Tam, A.W.; Smith, M.M.; Guerra, M.E.; Huang, C.C.; Bradley, D.W.; Fry, K.E.; Reyes,

Virology 185, 120-131, 1991

A:Title: Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length

A:Reference number: A40778; M01D:92024067

A:Accession: A40778

A:Molecule type: genomic RNA

A:Residues: 1-1693 <TAM>

A:Cross-references: GB:M32400; NID:q330023; P1DN:AAA5734.1; PID:q330024

R:Fry, K.E.; Tam, A.W.; Smith, M.M.; Kim, J.P.; Luk, K.C.; Young, L.M.; Platak, M.; F

Virus Genes 6, 173-185, 1992

A:Title: Hepatitis E virus (HEV): strain variation in the nonstructural gene region e

A:Reference number: A48547; M01D:92271462

A:Accession: A48547

A:Molecule type: genomic RNA

A:Residues: 967-1693 <PRY>

A:Cross-references: GB:M32400; NID:q330021; P1DN:AAA03206.1; PID:q330022

A:Note: sequence extracted from NCBI backbone (NCBI:104572, NCBI:104573)

C:Superfamily: hepatitis E virus nonstructural protein

C:Keywords: ATP; nonstructural protein; nucleotidyltransferase

Query Match	82.6%	Score 7448;	DB 1;	Length 1693;
Best Local Similarity	81.8%	Pred. No. 0;		
Matches 1396;	Conservative 110;	Mismatches 169;	Indels 32;	Gaps 6;
DB	1	PGITTAEOAALAAANSALANAVVVPFLSRVQTEILINMOPROVYFREVLMNHRTOR	60	
DB	10	PGITTAEOAALAAANSALANAVVVPFLSHQOETILINMOPROVYFREVLMNHRTOR	69	
QY	61	VINNELEOYCARAGRCLEYGAPHSINDPNVLRHCFRPGVDVORWYSAPTRGPAAN	120	
DB	70	VINNELELYCARAGRCLEYGAPHSINDPNVYHREFLRPGVDVORWYTAPTRGPAAN	129	
QY	121	CRSSALRGLPADRTYCEDFGFSRCAFAETGVALYSLHDMIPADVAMARHGXTRLYAA	180	
DB	130	CRSSALRGLPADRTYCEDFGFSRCAFAETGVALYSLHDMIPADVAMARHGXTRLYAA	189	
QY	181	LHPPEVLLPFGYHTSYLLIHGDRAVYTGSDTSAGYNHYSTRARITRTKIYGDH	240	
DB	190	LHPPEVLLPFGYHTSYLLIHGDRAVYTGSDTSAGYNHYSTRARITRTKIYGDH	249	
QY	241	PLVIERVRAIGCHFVLLTAAPESSPMYPVPRSTEVYRSIFGPGSGSLFPPSASCTK	300	
DB	250	PLVIERVRAIGCHFVLLTAAPESSPMYPVPRSTEVYRSIFGPGSGSLFPPSASCTK	309	

Qy	301	STHNVPHVIMDRLMFGATLDDOAFCCSLMUYLRGISKXVYGVLYVANEGBNMSDDL	360
Qy	301	STHNVPHVIMDRLMFGATLDDOAFCCSLMUYLRGISKXVYGVLYVANEGBNMSDDL	360
Db	310	STFHNVPHVIMDRLMFGATLDDOAFCCSRLMYLRGISKXVYGVLYVANEGBNMSDDL	369
Qy	361	TAXITAAVLTICHOVYLRTOAISKGMRLVEVNAOKFITLYLSWLEKSGRDIPEPRLQ	420
Db	370	TAVITAAVLTICHOVYLRTOAISKGMRLREHNAOKFITLYLSWLEKSGRDIPEPRLQ	429
Qy	421	FYAOCRMLSGFHLDPRLVFEDESVPCRCRTEPLKVAAGFCCFMRLQECTCELEPAE	480
Db	430	FYAOCRMLSGFHLDPRLVFEDESACHORTAIRKLSKFCCFMKMLQECTCELEPAE	489
Qy	481	GLVGHGHDNAYAGSGSEVDPAEPRLHLDVSTCYVHGOCLALYRALVNOODLAARSRLT	540
Db	490	GAVGDGHDNAYAGSGSDVDPEHISIDISGSYVPGALOPLYOALDPAEIVARRGRLT	549
Qy	541	ATVELVASPDLERFVYGVNKTFFTYVVDANHLAANGPEQYVLSFPAOSRMSGASHLT	600
Db	550	ATVKSQYDNRIDDEFTLLGNKTFPTSPYDVAVLETNGPERHNLSPFASOSYMAAGFSLT	609
Qy	601	YELPRLAGVRISSNGLDCTATPRLPGGASAPAEVAAFCALSALYRYNRFTQHSHTGLIM	660
Db	610	YAAASAGLEVRYVAAAGLDHRAVFAFGVSPRSASAGEVTAFCSALYRPNRACRHSHTGLIM	669
Qy	661	LHPCEGLIFPSPGHIWESANPFCEGRYLRTWS-TSGFSDSP-----PEANA	714
Db	670	FHPCEGLIFPSPGHIWESANPFCESTLYRTWSEVDAYSSPARPOLGFWSESPIS	729
Qy	715	MAATPGLPHSTPPYSDIWLVPPESEEPQVYAAAV-PPADPPAGLP-GPVVLTPPPPPVH	772
Db	730	RAAPRTL-----AALPPAPDPSPPPSAPALAPASATAG	766
Qy	773	KPSTP-PPSRKRLTYPRDQAKYVAGSLFESDCLWLVAANSNGHRPPGGCLCHAFYQRP	831
Db	767	APATTHQARRRRLFTFYPDGSKYFASLSESTCTMLVNASVNDHRPPGGCLCHAFYQRP	826
Qy	832	EAFPTERTIMEGSLAATLTPRPLTHAVADYVEQONPKRLAAYETCSRGCTAAYLL	891
Db	827	ASFDASATVMDGAATLTPRPLTHAVADYLEHNPKRLAAYETCSRGCTAAYLL	886
Qy	892	GGGIYOVVSLSPFAMERNHRPGDELTLTPRANWEKANPQAPVLTITEDARTANLAL	951
Db	887	GTGIYOVVIGSPFAMERNHRPGDELTLPLARWKEANRPRTPLTITEDARTANLAI	946
Qy	952	EIDATEYGRACAGCTISPGIVHIOFTAGVPSGSKRSIQGGDVVVVPTBELRNSWR	1011
Db	947	EIDNSTIDGRACACRYTPGVVOYOFAGVPSGSKRSITQADVDDVVVPTRELNNMR	1006
Qy	1012	RGFAPRPHHTAARTTORRVIDEABSLPHLLLLHMOAASSVHLLGDNQIPALDFEHA	1071
Db	1007	RGFAPRPHHTAARTTORRRVIDEABSLPHLLLLHMOARATVHLLGDNQIPALDFEHA	1066
Qy	1072	GLVPAIRPELAPTSMWAVYTRPCPADVCELLRGAYPKIQTTSRYLRSLEFNEPAIGOKLVX	1133
Db	1067	GLVPAIRPDLGPTSMWVHTRWRPADVCELLRGAYPMIQTTSRYLSLFPNGEPAVQOKLV	1126
Qy	1132	TQAAKANPGAIIVHEAOGATFTEHTTIATADARGLIOSSRAHAIVALTIRHEKCVILDA	1191
Db	1127	TOAAPPANPGSVVHEAOGATYFETHTTIATADARGLIOSSRAHAIVALTIRHEKCVILDA	1186
Qy	1192	PGLLRREGISLVYNNFFLAGGEVGHRSVPIRGPNDOGLTLOAFPPSCOSIAYHOLA	1251
Db	1187	PGLLRREGISDAIVNNFFLAGGEIGHORSPVIPRGPNDAVNDVTLAAPPSCOSIAYHOLA	1246
Qy	1252	EELGHRAPAAVLPCCPELEOGILVMPQELVSDSVLFEELTIDVHGMARPQSORKAVL	1311
Db	1247	EELGHRAPVAAVLPCCPELEOGILVLPQELTCDSVVFELTIDVHGMARPQSORKAVL	1306
Qy	1312	STLVGRGRKTKLYEAHSDVRESLARFTIPGVOATTCCELYELVEAWEVGOGSAYVL	1371
Db	1307	STLVGRGRGRKTKLYNASHSDVROSLARFIPAIGPVYTTCELYELVEAWEVGOGSAYVL	1366
Qy	1372	ELDLCNDSVSRITFPQKXCNKFTTGGETIAHGKVGOGISAMSKTFCALGPWFERALEKEIL	1431

Db	1367	ELDLCNRDVSNITFFQÖDCKNFTTGETIAHGKVGÖGISMSKTFPCALJFQPMFRALEKIL	1426
Qy	1432	ALLPNIIFYGAAVEESYFAAASGAGSCWFEENDFSEPDSTONNFSLECYVMECCMP	1491
Db	1427	ALLPGVGYGDAFPDIFYSAVAANAARMSVFEUNDFSEPDSTONNFSLEGECAIMBECMP	1486
Qy	1432	ÖMLIRLHYLVANSIILQAPKESLKGPKKHSGEPGLLNNTYNNNAI1AHCEEFDFRYA	1551
Db	1487	ÖMLIRLHYLIRSAIILQAPKESLKGPKKHSGEPGLLNNTYNNNAV1THC9DFDFOYA	1546
Qy	1552	AFKGGDSVYLCSVDYRÖSRNNAAL1IAGCGELKLYADYRPIGYAGVYVAPGLTGLTPVVRFA	1611
Db	1547	AFKGGDSVYLCSVEYRÖSPGAAYVLLAGCGELKLYADYRPIGYAGVYVAPGLTGLTPVVRFA	1606
Qy	1612	GRLSEKMNWPGPERAEQRLRLAVCDFLRLGTNNVAÖQVCDVYVSRYVGVSPGLVHNLIGLÖT	1671
Db	1607	GRLTFEKNWPGPERAEQRLRLAVSDFLRLGTNNVAÖQVCDVYVSRYVGVSPGLVHNLIGLÖA	1666
Qy	1672	IADGKAHFTETIKRVPDLNLSIÖRVE	1698
Db	1667	VADGKAHFTESVKRVPDLNLSIÖRVE	1693

RESULT 2

A44212 genome polyprotein - hepatitis E virus (strain Mexico)

N:Contains: RNA-directed RNA polymerase (EC 2.7.7.48)

C:Species: hepatitis E virus

C:Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 19-Jan-2001

C:Accession: A44212; B48547

R:Hung, C.C.; Nguyen, D.; Fernandez, J.; Yun, K.Y.; Fry, K.E.; Bradley, D.W.; Tam, A

VIROlogy 191, 550-558, 1992

A:Title: Molecular cloning and sequencing of the Mexico isolate of hepatitis E virus

A:Reference number: A44212; MUID:93079857

A:Accession: A44212

A:Molecule type: genomic RNA

A:Residues: 11691<HDA>

A:Cross-references: GB:M44506; NID:g330017; PIDN:AAA45730.1; PID:g330018

R:Fry, K.E.; Tam, A.W.; Smith, M.M.; Kim, J.P.; Luk, K.C.; Young, L.M.; Patak, M.; F

Virus Genes 6, 173-185, 1992

A:Title: Hepatitis E virus (HEV): strain variation in the nonstructural gene region e

A:Reference number: A48547; MUID:92271462

A:Accession: B48547

A:Molecule type: genomic RNA

A:Residues: 965-1691 <FR4>

A:Note: sequence extracted from NCBI backbone (NCBI:104576, NCBI:104578)

C:Superfamily: hepatitis E virus nonstructural protein

C:Keywords: ATP; GTP binding; nonstructural protein; nucleotide binding; nucleotidylt

F:973-960/Region: nucleotide-binding motif A (P-loop)

F:979/Binding site: ATP/GTP (Lys) #status predicted

Query Match	82.5%: Score 7433: DB 1: Length 1691:
Best Local Similarity	81.0%: Pred. No. 0:
Matches 1384: Conservative 121: Mismatches 166: Indels 38: Gaps	5:
QY 1 PGITTAIEQAAALAAANSALANAAVVVRPELSRVOTEILINLMQPROLVREPEVLMNPIQR 60	
Db 10 PGITTAIEQAAALAAANSALANAAVVVRPELSHQVEITLINLMQPROLVREPEVFMNPIQR 69	
QY 61 VHNHLEQYCARAGRCLEVGANPNSINDNPNVYLHRCFLRPVGRVQWYNSAPRTGPAAN 120	
Db 70 VHNHLEQYCARAGRCLEIGANPNSINDNPNVYLHRCFLHPVGRVQWYNSAPRTGPAAN 129	
QY 121 CRRSALRGLPADRFYCFDFGFSRCALFAAETGALTLSLHDLPADVAEAMARHGKXRLAA 180	
Db 130 CRRSALRGLPADRFYCFDFGFCGFAETGALTLSLHDLPADVAEAMARHGKXRLAA 189	
QY 181 LHLPEVLLPQGYHTTSTYLLIHGGRVAVVYEGDTSAGYNHDSILRAMITTKIYGDH 240	
Db 190 FHLPEVLLPQGYHTSTSYLLIHGGRVAVVYEGDTSAGYNHDAVLTLTWIRTKTVEEH 249	

```

QY 241 PVIETVRAIGCHFVLLTLTAPEPSMPYVYPRSTEVYVRSIFPGPGSPSLFSPASCTK 300
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 250 PVIETVRAIGCHFVLLTLTAPEPSMPYVYPRSTEVYVRSIFPGPGSPSLFSPACAVK 309
QY 301 STEHAVVHIDRLMFGATLDQAFCSSRLMTYLRGISYKVTVGALVANGMNASDAL 360
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 310 STEHAVVHIDRLMFGATLDQAFCSSRLMTYLRGISYKVTVGALVANGMNASDAL 369
QY 361 TAVTAAVLTICHOVRYLRTQAISGMRLGYEHQOKFTIRYSMLPEKSGDIYIPGRLO 420
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 370 TAVTAAVLTICHOVRYLRTQAISGMRLGYEHQOKFTIRYSMLPEKSGDIYIPGRLO 429
QY 421 FYACCRWMLSGFHLDPRLVLFEDSVPCRCRTFLKVAAGKCCMRMLGOCGCFLEPAE 480
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 430 FYACCRWMLSGFHLDPRLVLFEDSVPCRCRTFLKVAAGKCCMRMLGOCGCFLEPAE 489
QY 481 GLVGDHGHNEAYEGSEVDAPEPAHLDVSGTYAVHGHOLEALYRALNYPQDIAARSLIT 540
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 490 GLVGDHGHNEAYEGSEVDAPEPAHLDVSGTYAVHGHOLEALYRALNYPQDIAARSLIT 549
QY 541 ATVELVAPDRLECRIVLGNKTFRTYVDCAHLEANGPEQYVLSFDASROSGAGSHSLT 600
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 550 ATVELVAPDRLECRIVLGNKTFRTYVDCAHLEANGPEQYVLSFDASROSGAGSHSLT 609
QY 601 YELTPAGLOVRISNGIDCTATPPGAPSAPEVAAPFCALRYNRFTGRHSILTGLW 660
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 610 YELTPAGLOVRISNGIDCTATPPGAPSAPEVAAPFCALRYNRFTGRHSILTGLW 669
QY 661 LHPEGLIGIFPPSPGHIMESANFCEGLTYRTWTST-----SGFSSDSPPE 709
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 670 LHPEGLIGIFPPSPGHIMESANFCEGLTYRTWTSTIDTFLTVGLISG-HIDAPHS 728
QY 710 AAFAPMAATGCLPHSTPPVSDIWLPPPESEFQVDAAPVPAPDPLGPGVLTLPPPP 769
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 729 GGPAPATGTGAVGSSDSPDDP--PLPDVTGSRPSGAR-PGPPNPNGV----- 774
QY 770 PVHKHSTLPPSRNRLTYTPDGAKKVYAGSLFESDCDMLVNASPNRPGGLCHAFYQR 829
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 775 -----QRLHHTYPDGAKIYVSGTFESECWLVNASNAGHHPGGGLCHAFYQR 822
QY 830 FPEAFYPTERIMREGLAAYTLTPRPIIHAAVADYRVEDONPKRLAAARETSSRRGTAAP 889
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 823 YPSPFDATKTFVMRGLAAYTLTPRPIIHAAVADYRLEHNPKRLAAARETARSGTAAP 882
QY 890 LLSGSIYQVYVLSFDMERNHRPDELITLTPAANMFPAQPVLTITEDTARTANL 949
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 883 LLSGSIYQVYVLSFDMERNHRPDELITLTPAANMFPAQPVLTITEDTARTANL 942
QY 950 ALEIDATETVGRACAGCTISPGIYHYOFTAGVPSGSKRSIQCGDVVVVYPTRELNSW 1009
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 943 ALEIDSSSEVGRACAGCKVEGVRYOFTAGVPSGSKRSYQOADVYVVVYPTRELNSW 1002
QY 1010 RRRGFAFTPTTAAVYTGRRVVIDEAPSLPRLHLLHMORASSVHLLGDNPQIPADFE 1069
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1003 RRRGFAFTPTTAAVYTGRRVVIDEAPSLPRLHLLHMORASSVHLLGDNPQIPADFE 1062
QY 1070 HAGLVAIRBELAFTSMWXYVTHRCPADVCELIRGAYPKIOTTSVYLSLFWNEBAIQOKL 1129
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1063 HAGLVAIRBELAFTSMWXYVTHRCPADVCELIRGAYPKIOTTSVYLSLFWNEBAIQOKL 1122
QY 1130 VYVQAAAPAMPATVTHEAGATFTTTITATADARGLIOSSRAHAIVALTRHEKVCYL 1189
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1123 VYVQAAAPAMPATVTHEAGATFTTTITATADARGLIOSSRAHAIVALTRHEKVCYL 1182
QY 1190 DAGGLLEVGISDVIVNFFLAGEGVXHRPSVYPRGNPQONLGTLOAFPSCCISAYHO 1249
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1183 DAGGLLEVGISDVIVNFFLAGEGVXHRPSVYPRGNPQONLGTLOAFPSCCISAYHO 1242
QY 1250 LABELGRAPVAAVLPCELEOGLIYMPQOELTVSDSVLFEELTDIVHCHMAAPSQOKA 1309
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1243 LABELGRAPVAAVLPCELEOGLIYMPQOELTVSDSVLFEELTDIVHCHMAAPSQOKA 1302
QY 1310 VLSITVGRVGRRTKLYEAASDVRESLARIPTIGPVQATTCELVELVEANVEKQDQSA 1369

```

```

|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1303 VLSITVGRVGRRTKLYEAASDVRESLARIPTIGPVQATTCELVELVEANVEKQDQSA 1362
QY 1370 VLEIDLGNDRVSRITFEOKXCNKFTTGTTIAHGKVGQISAMSKTEFCALFGFWRAIKE 1429
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1363 VLEIDLGNDRVSRITFEOKXCNKFTTGTTIAHGKVGQISAMSKTEFCALFGFWRAIKE 1422
QY 1430 ILALTPIRIFEGDAVEEVPAAVAGASCWVFENDESEFPSTQNNFSLGLECYMEBCG 1489
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1423 ILALTPIRIFEGDAVEEVPAAVAGASCWVFENDESEFPSTQNNFSLGLECYMEBCG 1482
QY 1490 MPOMLIRLHYLVRSAWILQAKRESILKGFMKHSGEPGTLNNTVYMAALIAHCYEFDR 1549
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1483 MPOMLIRLHYLVRSAWILQAKRESILKGFMKHSGEPGTLNNTVYMAALIAHCYEFDR 1542
QY 1550 VAAFKGDDSVVLCSDYRQSRNAALIACGGLKLYKDYRPIGLYAGVVAAPGLTLPDVR 1609
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1543 VAAFKGDDSVVLCSDYRQSRNAALIACGGLKLYKDYRPIGLYAGVVAAPGLTLPDVR 1602
QY 1610 FAGRLSEKNMGPGPERAQLAVCDLRLTNNAQVVDVYSRYGVSPGLVHNLIGML 1669
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1603 FAGRLSEKNMGPGPERAQLAVCDLRLTNNAQVVDVYSRYGVSPGLVHNLIGML 1662
QY 1670 OTIADGRAHFTETIKPVLDLNTSIORVE 1698
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1663 OTIADGRAHFTETIKPVLDLNTSIORVE 1691

RESULT 3
C48547
nonstructural protein - hepatitis E virus (strain Tashkent) (fragment)
C:Species: hepatitis E virus
C:Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 20-Sep-1999
C:Accession: C48547
R:Fly, K.E.; Tam, A.W.; Smith, M.M.; Kim, J.P.; Luk, K.C.; Young, L.M.; Platak, M.;
Virus Genes 6, 173-185, 1992
A:Title: Hepatitis E virus (HEV): strain variation in the nonstructural gene region
A:Reference number: A48547; MUID:92271462
A:Accession: C48547
A:Molecule type: genomic RNA
A:Residues: 1-290 <FRY>
A:Cross-references: GB:110337; NID:q291457; PID:AAA4573.1; PID:q291458
A>Note: sequence extracted from NCBI backbone (NCBIN:104577, NCBIIP:104580)
C:Superfamily: hepatitis E virus nonstructural protein
C:Keywords: ATP; nonstructural protein

Query Match 15.2% Score 1368; DB 2; Length 290;
Best Local Similarity 85.9%; Pred. No. 1.6e-78;
Matches 249; Conservative 21; Mismatches 20; Indels 0; Gaps 0;

QY 1344 GPVQATTCELVELVEANVEKQDQSAVLEIDLGNDRVSRITFEOKXCNKFTTGTTIAHGK 1403
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 GPVQATTCELVELVEANVEKQDQSAVLEIDLGNDRVSRITFEOKXCNKFTTGTTIAHGK 60
QY 1404 VGGISAMSKTEFCALFGFWRAIKEILALLPNIIFYGDAVEEVPAAVAGASCWVFE 1463
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 VGGISAMSKTEFCALFGFWRAIKEILALLPNIIFYGDAVEEVPAAVAGASCWVFE 120
QY 1464 NDESEFPSTQNNFSLGLECYMEBCGMQMLIRLHYLVRSAWILQAKRESILKGFMKHSG 1523
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 NDESEFPSTQNNFSLGLECYMEBCGMQMLIRLHYLVRSAWILQAKRESILKGFMKHSG 180
QY 1524 EPGTLNNTVYMAALIAHCYEFDRVVAAPFGGDDSVVLCSDYRQSRNAALIACGGLK 1583
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 EPGTLNNTVYMAALIAHCYEFDRVVAAPFGGDDSVVLCSDYRQSRNAALIACGGLK 240
QY 1584 VDRPPIGLYAGVVAAPGIGTLPDVRVFAGRUSEKNMGPGPERAQLAV 1633
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 VGRFPIGLYAGVVAAPGIGTLPDVRVFAGRUSEKNMGPGPERAQLAV 290

RESULT 4

```


OY	1234	TLGFPPSCQISAHQIAEELGRRPARVAAVLPCELEBQGLLTPMPOELTVSDSVLYEL	1233
Db	1134	TM-----FATVPTKOLMNSLY-----	1151
OY	1294	TDIYHCMAAPSORKAVLSTLVGRYR-----RFTLYEAAHSDPREL	1336
Db	1152	---VHRNLELVSKTGYTMOEFYORCLPGNSFVLNDFVATMRIRDNFNIOPCRJTL	1208
OY	1337	ARFIPT-----IGVQATTC-----LYELVANEKODGSAYL-ELDLCNR	1378
Db	1209	SNLDPVALLIKNEQNLIVLTLACEBRRIPIGLENLVAMIKRMNPTDLAGYDITNM	1268
OY	1379	DVS-----RITFEQKXCKNFTTGETIANGKV-----	1405
Db	1269	SISIVDNFSEFVDEVLDLHLCVBRASSIQSFSDWFSCQPTSAVGOLANFENIDLPAFD	1328
OY	1406	-----QISMSKTFCALGEMFRAIEKEIILALPPN--I	1438
Db	1339	TYMHMIRKOPKSRLDNISOSEYPALOTIYHPRVVAAGVPKFKYLTFTLSMDSSKFE	1388
OY	1439	FYGDAYEASV--FAAAVSAGSGCMVEENDESPEDSONNESLGLCEVAMEEDGMQWLR	1436
Db	1389	FYTKKPEDIOEFSSDISJSDISEIIEILDCKSDKSQSFHPSIEALAIWEKIGDIDLIAW	1448
OY	1497	LYHLVRSAMILQAPKESLAK--FMKXHGSGPGILLMNT-----VMNNAIIAHCEYFDF	1548
Db	1449	MMSMGHKRTILODQOAKIKTLIYYQKSGDVTTFFIGNFTIINAACYASMLPLDKCK	1504
OY	1549	RYAAAFKDDSDVYLC--DYROSRRNAALIAAGCGIKLKVYDPRIGLYAGVVV--APGLG	1602
Db	1505	---ASFCCDDSLIYLPKGLEYPDIQIATYANLV--WNEPAKILFRKKYGJFCGKYIHHANGCI	1560
OY	1603	TLDPVYFAGLSEKMMGSPERAPBQOLRLACVDFLGLTINVA	1644
Db	1561	VYPPDLKLSKLNKL-VYEYVEEFRSLIDVAHSLENGA	1601

```

RESULT      6
S38480      nonstructural protein - rubella virus
C:Species:  rubella virus
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 26-Aug-1999
C:Accession: S38480
R:Gillam, S.
submitted to the EMBL Data Library, March 1993
A:Description: Nucleotide sequence of the nonstructural protein genes of rubella virus
A:Reference number: S38480
A:Accession: S38480
A:Status: preliminary
A:Molecule type: genomic RNA
A:Residues: 1-2115 <GTL>
A:Cross-references: EMBL:X72293; NID:g410507; PIDN:CA51087.1; PID:g410508
A:Superfamily: rubella virus nonstructural polyprotein

```

Query Match	4.3%;	Score 390.5;	DB 2;	Length 2115;
Best Local Similarity	19.3%;	Pred. No. 5.8e-16;		
Matches 452;	Conservative 205;	Mismatches 736;	Indels 947;	Gaps 105;

```

QY      3  ITTIEGALAAANSALANAVVPRFSRQTEILINLMQPROLVPRFVRLMHHPIQRVI  62
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db     40  VVTIAQKRAIV-----AVIPRPVFTQMOVS-----DHP--AL  69

      63  HNELOYCRARAGRCLEVGAPRS-----INDNENVLRGFLRPVGRDVRWTS-----  111
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db     70  H-ALSRYTRR--HWIEMG--PKALHLVLIDPSGL-----LREVAVERERWALCIHRT  118

      112 -----APTRGPAAN---CRRSALRGLEPP-----ADRTYC-FDGP  141
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db     119  ARKLATALAETASAWHADYVC---ALRGAPSGFYVHREDVPHGRGAAVADRCILLYTPM  175

      142  SRCFAAETGVALLYLHLDMPADVA-----E  167
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

```

Dd	176	OMELMTIDATLLVAVDMLPVALAAHVDDMDJLGMHLDHOGSPRADCSAGACGPPR	233
Oy	168	AMARHGTRLXAALHREPEVLLPRGTHTTSYLLIHDBGRAVTEYGDTSAGYNHDSIL	227
Dd	236	GTRPCPTRIYOVL---PDIAHGRRLXRCGRPLMTFDCAVAELSMEVACHGHQ-----	286
Oy	228	RAMIRTKIVGDHDLVIERVRAIGCHEVLLLTAAPESPMPVYVPRSTEVYVRSIFGBG	287
Dd	287	-----ARVRAVACTLPIRHVRSIQPS-----ARVR-----	311
Oy	288	GSPSLFBSAGCTKTFFAVYVNHIDRLMLFGALLDQACCCSLMYLTGISTYKTVAL	347
Dd	312	-LPDLVHLAEYGMWMPSLPRVFOHRLSYCKLSPDAYSEEVFKFNALSHSTLLGN	370
Oy	348	VANEGWN---ASEDALPAIXITAAVYLTCOHRYLTOAISKGMRLRGVEJAKOFTRLYSW	404
Dd	371	VLOGSKWGTCAEDALCAVYAFR---AMOSNRLAGIMKSKACABSLS---VAGW	421
Oy	405	LEFKSGRDYIPGROLOFYA-----OCRWLASGFHLDPRLVLFDESVPBCRGTEPLKRYA	458
Dd	422	L---DTIMDAIKRFFGGSVPLAEHMEW---EQDAVAVAFD-----	455
Oy	459	GKFECCPMWLOECTGLERABEGLVGDHGDNAIYAGSEVDAERAHLDVSGTYAVHGHQ	518
Dd	456	-----RGRLDGGRHLDTVQ-----PKSPRPRLAATFWIVHAAS	490
Oy	519	LE---ALYRALNVPOD-----RGLAEASRLT	540
Dd	491	ADRHCAACPRCDYRERPSAPACRPDDEALIPWLEAERALLRCRWDEALRAADTA	550
Oy	541	ATVELVASPDRLRECTVL-----GNKTRTYVDSGANELEANGPROOYLS	584
Dd	551	APAPLAPRAPAYR---TVLYHRPHAHNGFWLTDERGGADALVLCRPLQGRPLNGRENH--	605
Oy	585	FDASROSKGAGSHLTYELTPAGIQ---VRISNGLDCTATFPPGASAPARCEVAAFCSA	642
Dd	606	-----YAAGAHMCA---GARGLOAFVRV---VPRERPMADGG--ARAWAK	643
Oy	643	LYRYNRFLOHRSILGMLHREGLIG---IPPFSPGHMESANFCSEGLTYTRMST	698
Dd	644	FFRCGANAOR-----LGBEPAVMLPYTDGDPKLT-----ALALRTLAQ	683
Oy	699	SGFSSDS-----PEEAPARMAATPR-LPHSTPVSIDWILPRP-----SEBF	741
Dd	684	QGAALALSVROLPRGTAFFEAANVATAVVRAGQGLATSPRPD---PPRRARRRSQRH	739
Oy	742	QVDAAPVPRAP---DPAGLPCPVVLTGPPRPVHKPSIRP-----SNRRLLTYT--PD	791
Dd	740	SDAKGTPEPPAPVRDPR--RPOP---SEPAPRVGDV---VPTTAEAPDRAHAELEUVYERS	794
Oy	792	GAKYVAGSLPESD-----CDMLVNASPNHRRGGGLCHAF	826
Dd	795	GPPSTADPDSDIVESTYAAACGPVHLRYADIMDPPGCKVYVNAANEGLLAGSGVCGAI	854
Oy	827	YQREPEAFYPTETIMERGLAA---YTLTPR-----IHAUAV-----DY	863
Dd	855	FANMTAA-----LAADCRBLAPCIGEAVATPGHGGYTHIIHAUAVRRPRDPA	903
Oy	864	RVEONPKRLAEAYE-----TCSRGTAAVPLGSSIT---QYVPVISLSPDAMRRNHRPGDE	916
Dd	904	ALEGGALEERAYRISIVALAAARRMARVACPLGAGCYVQMSAAESIRALATTRAEPAR	963
Oy	917	LYL-----TEPAAMWFEANKPACQVLTITTDRTART	946
Dd	964	VSLHICHPDRATILTHASVLYGAGLAARRVSPPTPELAS--CPAGDPGRAPRQASAPATP	1022
Oy	947	ANALEIDAATEVGRACACTIS-----PGIV-----	973
Dd	1023	LG-----DAPAPERGCGGCELCRTRYRVTDRAVYVLMLEBRGATSMAMRIPEVVYGP	1077
Oy	974	-----HYQF-----TAGYVGSKSR-----	988
Dd	1078	EHLATHPLNHSYSLKAEPVPRPGMGCSMMHRCRGMQOMPOVRCPSNAHALCRLTGV	1137

```
QY 989 ---SIOGDDVVVPTRELIN-SMRRGFAAFTPHTAARTIGR----- 1029
| : : : |
Db 1138 PAVSTRGELDENCTCMLRAAANVAQAARACGAYTSAGCKAYGRALSERTHEDEPAALS 1197
QY 1030 -----RVVIDEAPSLPHLLLMQRA----- 1051
| : : : |
Db 1198 OMMSASHADASPDGTGDLPLMETVGCACSRVWVSGEHEAPPDHLVSLHRAHPGPGV 1257
QY 1052 -----SSVHL-----GDPNQI-----PAI----- 1066
| : : : |
Db 1258 VLEVAREPBGNTGTHFVAVGCGRRVSDRPHMLAVPLSGGTCATDEGLAQAYD 1317
QY 1067 DEE-----HAGLVPAIRPELAPIS--WXYVT----- 1095
| : : : |
Db 1318 DLEVRLGDADAMAARALASIQRRKGPYNIWVNMMAAGAKTTRILAAFTREDLYCPTN 1377
QY 1096 ----DYCELR-----GAYPKI-----QTSR 1113
| : : |
Db 1378 ALHHEIQAKLRARDIDIKNAATYERATKPLAAVRIYIDEAFTLGECYCAVASQTVAE 1437
QY 1114 VL-----RSLF--WNEP----- 1123
| : |
Db 1438 VICVGDROCCGPHYANNCTPYDPDKPPTGRSRHTFRFPDCAARLRAGLDYIEGERTGT 1497
QY 1124 -----AIGQKLVYQAAKAAKANGAITVHEOGATFETTTIATADARG- 1167
| : : : |
Db 1498 FACNLMDGROVDLHLAFSREYV--RRLHEAGIRATVREAGQMSVGTACIHNGROGTDVA 1555
QY 1168 IOSSAAHAIVALTRHEKCVIID-APGLIREVIGSDVYVNNFFLAGEV-----GXH 1218
| : : : |
Db 1556 LALTDLAIVLTRASDLYLHELEDGLLRAGLS-----AFLDAGALAELEKEVPAGID 1609
QY 1219 R-----PSVIRPKNPNQNLCTQAFPPSCQISAVHQLAEELGHAPAVAAVLPSCPRL 1271
| : : : |
Db 1610 RVAAVEQAPPLP---PADGIPEADQDVPPCPRTLEIEIVGKAGH---PHYVDLNRVTG 1663
QY 1272 EOGLLYPOEELTVSDSVVFEELTDIVHCRMAAPSOAKAVLSTLVGGRRTLYEAHSD 1331
| : : : |
Db 1664 EREVAYMR-----ISRHLNKNHTEMPTERTVLSAVSCGRAGE-----DGST 1707
QY 1332 VRESIARLP---TIGVQAT---CELYEYAMVEKQDGSVALEL----- 1373
| : : : |
Db 1708 LRTAAROHPRFPRIPPRPRTAGVAQEWKMTYLRERIDLTDTVYQMGVAARELIDRYTR 1767
QY 1374 -----DLGNRDSRIRTFQKXCKNF-----TTGETIANKGVGCGISAMSKTFCA 1417
| : : : |
Db 1768 RYPEIFAGCTAQSLSVPAFLKATLKCYDALGPRDTECHAAQGAAGLEIRAMAEWQ 1827
QY 1418 LFGPWFAIEKEIILALPPNIFYGDAYEESVFAAASGAGSCMFENDESEFSTQNNFS 1477
| : : : |
Db 1828 VMSPHFRALOKITMARLRQFLVAGHTEPEVDAMWQAHYTNALIEVDTEEDMNTLAT 1887
QY 1478 -----LGECVAMECCGPMQWILRLYLHVS--AMILQAPRESIKGFKKHSGEP 1525
| : : : |
Db 1888 RDVELEISAAALGLPCA--ED-----YRALRAGSYCTLRLEIGSETCERSCGP 1935
QY 1526 GTLLNNTYWNMAIILAHCEFR---DERVAA--FKGDDSVLSDYKROSNAA---ALI 1575
| : : : |
Db 1936 ATLENNTT---VAMCAMRNVPRKGVWAGIFQGDVNIPLPE--GAINAALKTWPAYV 1988
QY 1576 AGCGKLKLVY--RPIGLYAGVVAAPGLGLTDPVVFAGRLSEKMMGPREPEAEQRLAV 1633
| : : : |
Db 1989 GLFGHHIVYKHYSTPSPCG--HVGTAAGLFHDVHQAIVLCRRP--DPDVLBEQVAL 2045
QY 1634 CDLFLGLTNVAVCVDVSVRYGVSFGLVHNLIGMLQTIADGKAHETETIKVPLDTJNSI 1693
| : : : |
Db 2046 LDRLRG-----YYAALP-----DIVAANAAYXDYSAERVALIVREL 2081
```

```
RESULT 7
MMWVRN
nonstructural polyprotein - rubella virus (strain Therien)
```

N:Contains: nonstructural protein NS1; nonstructural protein NS2; nonstructural prote
C:Species: rubella virus
C:Date: 30-Sep-1989 #sequence_revision 30-Jun-1992 #text_change 16-Jul-1999
A:Accession: A35320; A29811
R:Dominguez, G.; Wang, C.Y.; Frey, T.K.
Virology 177, 225-238, 1990
A:Title: Sequence of the genome RNA of rubella virus: evidence for genetic rearrangem
A:Reference number: A35320; MUID:90281585
A:Accession: A35320
A:Molecule type: genomic RNA
A:Residues: 1-2205 <DOM>
A:Cross-references: GB:M15240; NID:9333971; PIDN:AA88528.1; PID:9333972
R:Frey, T.K.; Marti, L.D.
Gene 62, 85-99, 1988
A:Title: Sequence of the region coding for virion proteins C and E2 and the carboxy t
A:Reference number: A29811; MUID:88226020
A:Accession: A29811
A:Molecule type: genomic RNA
A:Residues: 1737-2205 <FRE>
A:Cross-references: GB:M15240
C:Comment: The cleavage sites of this polyprotein have not been determined.
C:Superfamily: rubella virus nonstructural polyprotein
C:Keywords: nonstructural protein

Query Match 4.2%; Score 374.5; DB 1; Length 2205;
Best Local Similarity 19.2%; Pred. No. 6,3e-15;
Matches 454; Conservative 201; Mismatches 714; Indels 995; Gaps 104;

```
QY 3 ITTAIEQALAAANSALNAVAVVPRFLSKVQTEIILINMOPROLVPRVYLNHPQRIY 62
| : : : |
Db 40 VYTAQKRAIV-----AVIPRVETQMOVS-----DHP---AL 69
QY 63 HNELEQYCARAGRCLEVGANPRS-----INDPNVILHRCFLRPVGDYQWYS----- 111
| : : : |
Db 70 H-AISRTYR---HWIEWG--PREALHYLIDSPGL-----IREVARKRYVALCLHRT 118
QY 112 -----APTRGPAAN-----GRRSALRGLPP-----ADRTYC-FDGF 141
| : : : |
Db 119 ARKLATALLETASSEAMHAYVC---ALRGAPSGPYVHREDVPHGCAVADRCLLTYTDM 175
QY 142 SRCAFAETGVALYSIHDIAPADVA-----E 167
| : : : |
Db 176 QMCLEMTITDAITLVAVDMPVALAHVGDMDMDLGIAMHLHDGSCPADCRGAGAPTP 235
QY 168 AARHGXTRLYALAHLPREVLRLPRGYHTTSYLLIHNDGRVAVTYGDSAGYNHDVSL 227
| : : : |
Db 236 GYTRPCTTYIYVL---PRTAHGRLYRCGPRLMTWDCVAELSWEVAAOCHGQ----- 286
QY 228 RAMIRTTKIVGHPVLEVRRAIGCHFVLTLTAAPERSPMPYVVPYRSTEVYVRSIFGPG 287
| : : : |
Db 287 -----AVRAVRCITLPIRHVRSIQPS-----ARVR----- 311
QY 288 GSPSLPSPSACSKSTFHAVPVHIMDRMLFGATLDDQAFCSRLMYTLNGISYKYTVGL 347
| : : : |
Db 312 -LPDVLHLEAVGRWRFSLPRPVFORMLSYCTLSPDAYSEVFEFKNALCHSTLACN 370
QY 348 VANEGWN---ASEDALTAKITAAVYLITGHQVRLQALISKGMRRIGVEHAOKETIRLVS 403
| : : : |
Db 371 VLQEGKGTCAEDALCA-----VFAFRAMOSNARLAIMGAKCAADSLSV 417
QY 404 ---WLFKSGRDYIPGROLOFYAACRRMLSA-----GFHLDPRVLVDESPCRRTF 453
| : : : |
Db 418 AGML-----DTI-----WDAIKRFLGSVPLAEHMEEMEDDAVAAFD----- 454
QY 454 LKKVAGKFCCEFMKWLQGECTGLEPAEGLVGDHGDNEAYEGSEVDPAEPAILDVSGTYA 513
| : : : |
Db 455 -----RGPLEDGGRIHDTVQ-----PPKSPRPPEIAATWI 484
QY 514 VGHQLE---ALYRALNVPQD-----IAARS 537
| : : : |
Db 485 VHAASEDRHCACAPRCVDYRERPPSAPAGPDDEALIPMLFAERRALRCREWDFFELRAR 544
```

QY	538	RLTATVELVASPDRLBERRVL-----GKTRTIVVDGALHEANGPEOY	581
Db	545	ADTAAADAPPARPRARYPTVLXRNPHNHGSPWLTLDPEGDAALVLCBPLGQRLPERH	604
QY	582	VLSFDASROSMGAGSHSLFTVELTPAGLO--VRISNGLDCTATPPPGAPSAAGVEVAF	639
Db	605	-----FAAGAHMCA--QARLOLQFVRY-----VPPERPAADG--ARA	639
QY	640	CSALRYNRFTQHRSLTGGMLHPEGLG----IFPPSPGHIMESANPFCGECTLYTRT	695
Db	640	WAKFFRCACMAOR-----LLEGPAVMHLPTDGDVPQLI-----ALALRT	679
QY	696	WSTSGFSDDS-----PPEAAAPAMAT-----PGLP-----HS	724
Db	680	LAOGGALALSVBDLPGCAFDANAVTAARVAGPQOASAPPPDDPPRRARRSORHS	739
QY	725	-----TPPVSDIWLPPPESEFQVDAPVPPADPAGLPGPVLTGTPPPPVHK-----	773
Db	740	DARCTPPPAARPPPP-----APSPAPPPAGDPP--PIPGAPDRARDAELE	787
QY	774	----PSTIPPSRNR-----RLLYTPDGAKVYAGSLFSDCDML	808
Db	788	VACEPSGPSTRADEPDSDIVESYARAAPVHLRVRDIDMPPGCKV-----V	835
QY	809	VNASNPCHRGGLCHAFQRFPEARUYPTPEIFMREGIA--YTILTPP-----	854
Db	836	VNAHEBELLAGSVCGLATFANATAA-----LANCRRLAPCPTGAVALTPGHGC	884
QY	855	-----IHAIVAP-----DYREONKPKLEAAYRE-----TCSRGTAAAPLLGSGIY--QV	898
Db	885	GYTHIIHAIVAPRRPROPAPALEBEGALLEEAYISYALAAARMACVACPRLGAGYTGSA	944
QY	899	PVLSLFDAMERNHRPDELYL-----TEPANMFE	928
Db	945	AESLRALATATPREPERVSLATICHDPDRATLTHASVYAGLAARVSPPTPELAS--CP	1003
QY	929	ANKRAQVULTTEDTARTALALEIDALEVGRACAGCTIS-----	969
Db	1004	AGDGPRAQBSASPPATPLG-----DATAPRPGCGGELCXYTRVTNDRAVYVNLMLERD	1058
QY	970	-----PGIY-----HYOF-----	977
Db	1059	RGATSMAMRIPEVVYVGPPEHLATHPFLNHTSVLKALEVAPRPMGCGSDMRCRGHNHQPQ	1118
QY	978	-----TAGVPSGSKRSIQOGSDVDVVVYPTRELSN--SMRRRGPAFTPHATA	1023
Db	1119	VRCTPSNAHAALCRTOGVPRASTRG--GELDPNOCMLAANVANVAQARACAVYTSACP	1175
QY	1024	RVTIGR-----RVVIDEAP	1037
Db	1176	KCAVAGRLSEARTHEDEFAALSORMSASHADSPDGTGPDLPIMETVGCACSRVWVGSEH	1235
QY	1038	SLPRLHLLLMHORA-----SSVHLL-----GPPNOI-----PAI-----	1066
Db	1236	EAPRDHLIVLSLHRAPNGPWCYVLEVARAREGNGPTGHEFCAGVGGPKRKSDBRPHLMTAVP	1295
QY	1067	-----DPE-----HAGLVPAIRPLAFTS--WNXYT--	1090
Db	1296	LSRGGTCATDGLAQAYVDYDLEVRRLDDDMARALALASVGRPKGYNIRVYMMAMAGA	1355
QY	1091	-----HRCRA-----DYCELIR-----GAYPKI--	1108
Db	1356	GKTRILIAATFREDLVCPFNALLHEIOAKRLARIDIKNATYERRILTKPLAAVARIYI	1415
QY	1109	-----QTSRVL-----BSLF--WNRP--	1123
Db	1416	DEAFTLGGECYCAFYASOTTAETVLCVGDROCCGPHYANNCRTPEVPDRMPTEKRSRHTWRPD	1475
QY	1124	-----AIGOKLVYTOAKKANPAGATVHE	1147
Db	1476	CMARLALAGLDYDIEGRCTCFACNMIDGROVDLHLSAREYV--RRLHEAGIRATYATRE	1533
QY	1148	AQGAFTETTTIATADARGL--IOSSRAHAIVALTTRHTEKCVILD--APGLREVGISDVIY	1205

```

Db      1534 AOGKMSVGTACIHVGKRGDIDVALATLRDLATYSLTRSDSLYTHELEDGLRAGLS---- 1589
               ||| : | : : ||| : : : | | |
Oy      1206 NNFFLAGEEV-----GXNR-----PSYIPGPNPONLGTLQAFPPSCQISAYHOL 1250
               ||| : | : : | | |
Db      1590 -AFLDAGALMLKEVPAGIDRVAVAEQAPRLP---PADGIPEADVDVPRPCRTLEELV 1644
               ||| : | : : | | |
Oy      1251 AEELGRPAVAVALPCEPELOGLLYMPQELTVSDSVLFVEFLTDIVHCMAAPSORKAV 1310
               ||| : | : : | | |
Db      1645 FGRAGH---PHYADINRWTEGEREVRVM---RISRHLINKNHTEM-----PGTER-V 1689
               ||| : | : : | | |
Oy      1311 LSTLVGRGRKKLKLEAAH--SDVRESLARFRP---TGPRQATP-----CETVEL 1356
               ||| : | : : | | |
Db      1690 LSAVA-----VRRIRAGEDGSTLRTAVARQHPRPRQIIPRPVITAGVAOEWRMYTLRER 1744
               ||| : | : : | | |
Oy      1357 VEAMVEKGODGSASVIEL-----DLCNDVSRIPTFEOKXCNKF-----T 1394
               ||| : | : : | | |
Db      1745 IDLTDTVTQMGVAHAELDRVARRYREIFAGMCTAQSSVPAFLKATLKCVDAALGPRT 1804
               ||| : | : : | | |
Oy      1395 TGETTAGHGVCQGICSAWSKTFCALFGPMFRAIEKETILLPNNIFYGGAYESVFRAAVS 1454
               ||| : | : : | | |
Db      1805 EDCHAAGACAGLEIRAMAKEWQVMSPHRAIOKIIMRALROFLVAAGHPTEVDAMNQ 1864
               ||| : | : : | | |
Oy      1455 GAGSGCMVENEDESEEDSPSQNNFS-----IGLECVMNEECGMPLRLYLVS 1503
               ||| : | : : | | |
Db      1865 AHYTNAIEVDTEDEDMNTLATRDVELEISAALLGPLCA--ED-----YRALTA 1912
               ||| : | : : | | |
Oy      1504 -AMITQAEPRESIKGFWMKHSGEGTLMTVMNMALIIACHCYFR--DFRYAA-FKGD 1557
               ||| : | : : | | |
Db      1913 GSCTCLRELGTSETGCERSGEPATLLNHT-----VANCMAMRMVKPVKNAGIFQSGD 1967
               ||| : | : : | | |
Oy      1558 SVYLCSDYRKQSNAAL-----IAGCGKLKYDRPIGLYAGVVAAPGLGTPDVVR 1609
               ||| : | : : | | |
Db      1968 WFIPEPARS---AALKWTPAEVLGFENHPVKNHSFTPPSCG-HVGTAGLFDHDVV 2023
               ||| : | : : | | |
Oy      1610 FAGRSEKNMGSGPPERAREQLRLAYODFLRGLTNVAQVCDDVVSRYGVSPGLVHNILGL 1669
               ||| : | : : | | |
Db      2024 QAIKVLCRFE--DPVPLEEQVALLDRLG-----YVALP----- 2057
               ||| : | : : | | |
Oy      1670 CTIADKAHFETTIRKPVLDLTNSI 1693
               ||| : | : : | | |
Db      2058 DTVAANAAVYDYSAERVLAIVREL 2081
               ||| : | : : | | |

RESULT      8
A38196
non-structural hypervariable probable nucleotide binding protein - hepatitis E virus
C:Species: hepatitis E virus
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 28-Feb-1997
C:Accession: A38196
RTsarev, S.A.; Emerson, S.U.; Reyes, G.R.; Tsareva, T.S.; Legters, L.J.; Malik, I.A
Proc. Natl. Acad. Sci. U.S.A. 89, 559-563, 1992
A:Title: Characterization of a prototype strain of hepatitis E virus.
A:Reference number: A38196; MUID:92115700
A:Accession: A38196
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-141 <TSA>
A:Experimental source: strain SAR-55
A>Note: Sequence extracted from NCBI backbone (NCBIN:75998, NCBIP:76003)
C:Superfamily: hepatitis E virus nonstructural protein
C:Keywords: ATP

Query Match          4.1%; Score 372; DB 2; Length 141;
Best Local Similarity 50.3%; Pred. No. 1.7e-16;
Matches    84; Conservative   12; Mismatches   29; Indels    42; Gaps     7;

Oy      651 ORHSJLGGJWLMEPEGJLGFPPSPGHIMESANPCGEGCTLYTRTMS-----TS 699
               ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db       2  QRLSLTGNFWFPBGLGFAPFSFGHWESANPCGEGCTLYTRTMS-EVDNAVSPAPDPL 61
               ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Oy      700 GRSSDFSPEEAAPAATAATPGLPHSTPPVSDTWLVLPPESEEQOVAAPV-PPAPPA-GL 757
```

Db 62 GFPS-----EFSIPSRAP-----TP-----AAPLEPPAPDPSPTL 93
OY 758 PCGVVLTLPPEPPVHKHSIP-PPSRNRRLTYTPDCAKYAGSLFSS 803
Db 94 SAPAGEPAPGATAPARALITHOTARHRLLETPDGSKYVAGSLFSS 140
RESULT 9
S01956
hypothetical protein, 195k - turnip yellow mosaic virus
C:Species: turnip yellow mosaic virus, TYMV
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 18-Jun-1999
C:Accession: S01956
R:Morch, M.D.; Boyer, J.C.; Haenni, A.L.
Nucleic Acids Res 16, 6157-6173, 1988
A:Title: Overlapping open reading frames revealed by complete nucleotide sequencing of
A:Reference number: S01955; M01D:88289359
A:Accession: S01956
A:Status: preliminary
A:Molecule type: genomic RNA
A:Residues: 1-1844 <MOR>
A:Cross-References: EMBL:X07441; NID:962222; PIDN:CAA30322.1; PID:962224
C:Superfamily: egplant mosaic virus RNA-directed RNA polymerase
Query Match 3.9%; Score 349; DB 2; Length 1844;
Best Local Similarity 20.6%; Pred No. 1.9e-13;
Matches 385; Conservative 203; Mismatches 649; Indels 634; Gaps 93;
OY 157 LMD---LMPADYAEAMARH-GXRLYALHLPEVLLPFG-----TYHTSYLLIHD 204
Db 141 MMDALMYHPSQIMDLFLKRPNLERLYASLVPPREALSDQSFPKLYTTRHTLHY- 199
OY 205 GBRVVTYGGDSAGYNHVSILKAMITRTKT-VGDHPLYTEKYALG-CHFYLLLTAP 262
Db 200 ---VPGHENGASYNP-SDASWLRINSIRLGNHLSVTILESGVPYHSLIORGP 252
OY 263 EPSMPYVYPRSTEVYVSIFGP----- 286
Db 253 PDPSPLOAPPTLMTSLDFSYQYPRLDVVSFRIPDAIELPQATFLOQPLRDLVPPAVN 312
OY 287 -----GGSPSLFSPSACSTKSTFHAVVHIMDLMLFGATLDDQAFCCSRLMT 333
Db 313 ALFTYRAVRLTSDPAFAFVRHMSKRPDHVWTSNAMDVLQTF-----ALLNV 361
OY 334 YLR-GISYVTVG-----ALVANEGMNASEDLTAKITAYLTICQRYIKRTQATISKMR 387
Db 362 PLRPVNVYVLOSPLASLSLTLKQHRRL--TATAVPLISFTLL-QRFV----- 408
OY 388 RLGEVHAQ-KETTRLYSWLFEKSGRDYIPGROLQFYAOCRRWLSAGFHLDPRVLVDESY 446
Db 409 PLRPLAEVKSITAFRELRYRKE-----FHHPLDV-----FHLOHRKRWHSAL 453
OY 447 PCRCRTFLKVAAGKCCFMRMLGQECTCFLPEAGLVGDHNDNEAYGEVDPAPAH 506
Db 454 ----- 511
OY 507 DVSCTYAVHGHOLEALYRALNVPQDIAARA-----SRLATYELVASPR-- 551
Db 465 KLP-----HALQKALLLRIPSLTATPPFERSEQSKMLPNALSLWTLKRFALPQASL 519
OY 552 -----LECRVLGKTRTYVDGAHLAANGPDEQVLSFDASROSAGSHSLYELTPAG 607
Db 520 VLLALSESSILHKLFSPPLQAH-----DTYHNLHPGSLSIOWERTP-- 564
OY 608 LQVRISNSGL---DCTATPPGAGPAPSAAPGEVAAFCSALYRNRRTQORSLTGLMLHPE 664
Db 565 LSIPTTALPPTPTSTAPDSEASLP---PAFAST-----FVPR----- 603
OY 665 GLLGTFPP--FSPG---HWESANPFCGEGTLTYRTWSTSGF-----S 702
Db 604 -----PPPAASSPQAQPPTTTAAAPTLEPT--QRTHONSDLALESSTSTPPPPIRS 655

OY 703 SDFSP-----PEAAAPAMAATPGL--PHSTPPVSDIWLPPSPSE----- 740
Db 656 PDMTPSAPVLPFLINSRRRRPQLPATPDLEPAHTPPPLISIPQDDPSADMLGSHLLH 715
OY 741 FOYDAAPVPPAPDPAGLPCEVVLTPPPPPV-----HKPS-- 775
Db 716 HSLPAPPTPLPSSQLLPALVLTNDPTAIGVLFEEELHPRRYEDNATFLTRLRLSLSNMH 775
OY 776 IPPSRNRRLTYTPDCAKYAGSLFSSDQDLVNA--SNPGRHPPG-----GLCH-- 824
Db 776 LPQPTLN-CLLSAVSDQFKVSEELHWSLQTLIPDSOLSSEETNTLGLSTHLLTALHLY 834
OY 825 -----AFYGRFPAFYPTFERIMREGLAAYLTFRPIIHAAVPDYVEONKRLAEAYRET 879
Db 835 NPGATVYSDRGPIILFGPSDITK-----IDITHTGPPSHNSPG-KRLIGSP-- 881
OY 880 CSRRGTAAVPLLG-----SGLYOVVLSFDAMERNHRPGDELY--LTEPAAMFEA 929
Db 882 -SAKGHPDPLIRAMKSFVSGNY-LPFS-----EAHNPSTISHAKNLISMKNKGFQ 933
OY 930 NKPAQVLTTEBTATANALEIDATEVGRACACTISPC-----IYHQTAGVPSSG 985
Db 934 VLSLDVSTGQRTGPPKERRIQLIDHYLDT-----NPGKTPVYH--AGFAGCG 981
OY 986 KSRSIQO-----GDVVVYVPTRELNSMR-----RGEFA--FTPHTAARYTIGRRV 1032
Db 982 KTYPTOOLKTKLFKPR-VSCETTELRTMKRTAMELHGSQSRFNTWESSILKSRILY 1040
OY 1033 IDEAPSLPRLHLLHMQRASSVHL--LQDP-----NOIPAIDEHAGLV 1075
Db 1041 IDEIYKMPKRYLDLSILADALELVIILGDPLOGEYHSGKSSNNHRLPS--ETRLRLP 1097
OY 1076 AIRPELAPTSWMXVTHRCRADVCEL--RCAYPRIQ-----TSRVLRSLF 1119
Db 1098 YI---DMTCWM--STRIPQCLARLPQIHSFNAMOGVIGSVTPHDQSPVLTSHASSLT 1151
OY 1120 WNEPAIGOKLVYTOAKAANPGATFVHEAGATFETTTIATDANGLIQSSRAHIVAL 1179
Db 1152 FN-SIGYR-----SCTISSOGGLFPCPAILVLDNRYKWLSS--ANGVAL 1194
OY 1180 TR-----HTEKCYILDA--POLREVGISDIYVNNFL 1210
Db 1195 TRSRSGVQFMPGSPSYGVGTNGSSAMFSDAFNNSLIIMDRFPSPFLQFL--KLTSPSLTT 1251
OY 1211 AGEVGVKHRPS-----VIRGNPDQNLGLOAFPPS 1241
Db 1252 RGPRLNGATPPSASPTHRSPFHLRPHIPLSYDRDEYTVNPTLPDQGETRDLT--HFLRP 1309
OY 1242 QOISAVHQLAEELGHRPAPVAAVLPCCPELEQGLLYMPOELTVSDSLVLEL-----T 1294
Db 1310 SRLPLHFDLPAL--TPPPYSTVDP-PQAKAPVY-PGAF--FDSLAAFFLPAHDPSTR 1363
OY 1295 DIVHCEMAAPS---ORKAVLSTLVGRYGRRTKLYEAAH-----SDVRESLARPT 1340
Db 1364 EILHKDOSSNQPFWRPESLSC-----QPSLLISAKHAPNHDPTLLPASIKRR--RFR 1416
OY 1341 PTIGPYOATP-----C-----ELYELVAMEYKGGDSAYLE 1372
Db 1417 PSDSPHOITADADVVLGLQLFHSICRAYSRQPNSTVPPNBELEAGTSLNLEYAOLSKTOS 1476
OY 1373 LDCLNQDV-----RITFEOKXCKFTTGETIAGKYQOGISAMSKTFCALGFP--W 1422
Db 1477 TIVANASRDQPMWRHTTVKIFAAQKRVNDGSLFGSMKACQTLALMHDVYLVLGVYKKY 1536
OY 1423 FRAIEKEILLAPPNIF-----YGADESVFPAAYVSGAGSCMVENDFSEPDSTQ 1473
Db 1537 ORIFDN--ADRPNIYSHCGKTPNQLRDMCOEHL-----THSTPKIANDYTADDOO 1586
OY 1474 NNFSLGLECVVMEGMPQMLIRLY-HIVRSANITLQAPKESLKGFMKKHSGEGTLMMNT 1532
Db 1587 HGESVVALKMKRLNIPSHLIQHLVHLKTNVSTQPGPLTCM-----RLTGEPTYDDNT 1641

QY 650 TORHSLTGLMLHEGLLGFPPSPGHIMESANPFCGEGTLYTRWTSJGFSFSDPSPE 709
Db 622 -----LQPTTSKEBSLKT-----VSDVGESSIKE 648
QY 710 AAAPAMAATPGLPHSTPPVSDIWLPPSPSEFVDAVPAPDAPGLPQPVLLTPPPP 769
Db 649 VVKRSEISMGLGNT--VSEDEFORSTEIESLO-----679
QY 770 PVKRPSTIPPPSRNRRLIYTPDGAKYVAGSLFESDC---DWLVNASNPGHPPGGGLCHA 825
Db 680 QFHMVSTETIIRKOM-----HAMVYTGPLKYQCKNYKNDSLV-----746
QY 826 FYQRPPEAFPTPEFIMEGLAATLTPRPIIHAVAPDYREQNPRLAAYRETCSRGT 885
Db 717 -----ASLSAAVSNLKLTKDTA--AIDLETKERGVY-DVCLKWL 755
QY 886 A-----AYPLGSGIYQVPSLSPDAMERNRHPGDELYTEPAANWFANPQAQVLT 938
Db 756 VKPLSKGHANGVMDSDYKCFVAL-----LYDGEN-----I 787
QY 939 ITEDTARTANLALEIDAATEVGRACAGCTI---SPGIYHYQFTA--GVPGSGKSRSIQ 992
Db 788 VCGETWRRAVAVSSESLVYSDMGKIRAIRSVLKDGEPHISSAKVTLVDGVPGCKTKELLS 847
QY 993 G---DYDVVVVPTRELNSRRR---GFAFTPHTAARY-----TIGR-----RVID 1034
Db 848 RVNFDDELVLVPGQAEMERARRANSGLVAATKENVRTVDSFLMNGRPGCCYKRRFLD 907
QY 1035 EAPSLPHLL--LHMQRASSVHLGDPNOIPAID---FEHAGLVPAIRBELAPTSWX 1088
Db 908 EGLMLHPCGNFLVGSLSCEAFVYDQOIPYINRYATPPYKHLSOLEVDAVETR--R 965
QY 1089 VTHRCPADVCELIRGAYP-KIQTTSVRLSLFW-----NEPAIGOKLVYTOA 1135
Db 966 TLRCPADITFFLNQKVEGVMCTSSVTRSVSHEVLOGAAMNPVSKPLKGVITFTQSD 1025
QY 1136 KA-----ANGATVHEAGATFTETIT-ATADARGLIOSSRAHAIVALTRHEK----1185
Db 1026 KSLLSRGYEDVTHVHEVOGETEDVSLVRLPTPVGIISKOSPHLVSLSRTRRSIKYY 1085
QY 1186 CVLLDA--PGLLREVG--ISDIYVN-----NFLAGGEVGHHRPSY 1223
Db 1086 TVVLDNAVSVLRDECVSSYLDMKVDVSTQOLOIESYKGVNLVFAAPKTS-----1138
QY 1224 PRGNPDNLGTLQAFPPSQ-----ISAYHOLAEELGHRPAPVAALPPCPLEOGLLY 1277
Db 1139 -----DVSMDQYDYDKCLPGNSTILNEYDAYTMQI-----1168
QY 1278 MQEELTVSDSYLVF-----ELTDIYHCHMAAPSQR--AVLSTLVGRYGR--1321
Db 1169 RENSILNVKDCVLDMSKSVPLPRESETTLKRVITAEKPRKPGLLLENLVAAMIKRNFNSPE 1228
QY 1322 -----TKLVEAHSADVRESLARFIPTIGVQATTCCLYELVEAMVEKGOD 1366
Db 1229 LVGVVDIEDTASLYVDKFD--AYLKEKKKPKNIPILSRAS-----LEKWIKE--OE 1277
QY 1367 GSAVLEL-----DLGNRVSRITFPQKXNKFTTGETTAHGKVGOGISAMSKTECALFG 1420
Db 1278 KSTIGOLADFDLIDPRAVQYRHMIKQPKORLDLSIQTEYRAL-QTIYVHKKKINALFG 1336
QY 1421 PWFRAIEKEIILALPPN--IFYGDAVEESY--FAAIVSAGSCMWFENDPSEFDSTQNNF 1476
Db 1337 PVFSELTROLLETIISSRPMFYTRKPTQIEEFPSDLSNVPMDIILEDISKYDKSQNEF 1396
QY 1477 SLGLECVNVEEGCMQWMLRLHIVRSAMILQAPRESLKG-----WKHSGEPGT 1527
Db 1397 HCAVEYEIKRIGLDLEAV-----WKHGRKTTLLKDYTAGIKTCLMYQKSGDVTT 1449
QY 1528 LLMTVWMMNAITAHCEYFDFRVAAFKGDSDVYLQ---SDYROSRNAALIAGCGLKLKY 1584
Db 1450 FIGNITIIAACSMLPMERLLKGAFCGDDSLIFPKGTDFPDIOGANLL--WNEFAKL 1507

QY 1585 DYRPIGLYAGVVA-----PGIGTLPDVYVFRAGRLSEKNNMGGPPERRAQLRLAVCDFLRGLT 1641
Db 1508 FRRRYGFCGRYIIHHDRGCIYVYDPLKLISKLGAKHI-KNRHLEEFRTSLCDVAGSLN 1566
QY 1642 NVAOV--CVDVVSRYVGVSP 1659
Db 1567 NCAYTYTHNDVAGEVITKTAP 1586

Search completed: March 7, 2002, 14:04:18
Job time: 141 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2002, 14:03:52 ; Search time 16.68 seconds
(without alignments)
3732.427 Million cell updates/sec

Title: US-09-468-147-91

Perfect score: 9013

Sequence: 1 PGTTTAEQALAAANSALA.....FTETIKPYLDLTNSIQHVE 1698

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7459	82.8	1693	P33424	hepatitis e
2	7448	82.6	1693	P29324	hepatitis e
3	7433	82.5	1691	P03495	hepatitis e
4	7433	82.5	1693	P01HEVME	hepatitis e
5	406.5	4.5	1648	P19523	cucumber gr
6	374.5	4.2	2205	P13889	rubella vir
7	351.5	3.9	1612	P08433	odontogloss
8	349	3.8	1844	P10358	turnip yell
9	346	3.8	1844	P28477	turnip yell
10	339.5	3.8	1612	P86659	odontogloss
11	321.5	3.6	1616	P03387	tomato mosa
12	321.5	3.6	1616	P09Y46	tomato mosa
13	320.5	3.6	1616	P09Y46	tomato mosa
14	320	3.6	1597	P06220	chinese rap
15	319.5	3.5	1616	P90211	tomato virus
16	316.5	3.5	1616	P89676	tomato mosa
17	313	3.5	1601	P08820	turnip yell
18	312.5	3.5	1612	P89657	pepper mild
19	312.5	3.5	1612	P29098	pepper mild
20	311	3.5	1748	P35228	erysimum la
21	310	3.4	1844	P20128	turnip yell
22	305.5	3.4	1776	P20127	onion yell
23	301.5	3.3	1616	P09308	tomato mosa
24	299.5	3.3	1616	P09Y46	tomato mosa
25	295.5	3.3	1616	P03386	tomato mosa
26	293.5	3.3	1616	P30738	tomato mosa
27	293	3.3	1874	P36304	kennedy ye
28	291.5	3.2	1616	P09874	tomato mosa
29	287.5	3.2	1839	P20126	eggplant mo
30	261.5	2.9	1609	P18339	tomato mll
31	254	2.8	1629	P89202	sunh-hemp m
32	243.5	2.7	1707	P05080	tobacco rat
33	243.5	2.7	1884	P27738	apple chlor

34	242.5	2.7	1885	1	RRPO_ACTLSA	P54891	apple chlor
35	218	2.4	2512	1	POLN_STNDV	P03317	sindbis vir
36	211	2.3	2514	1	POLN_STNDV	P27283	sindbis vir
37	207	2.3	2514	1	POLN_ONNVG	P13886	O'nyong-nyo
38	201.5	2.2	1718	1	RRPO_SHVX	004575	shallot vir
39	200	2.2	2431	1	POLN_SEV	P08411	semliki for
40	194	2.2	1967	1	RRPO_PVMR	P17965	potato viru
41	194	2.2	2485	1	POLN_PEVV3	P36327	venezuelan
42	184.5	2.0	3175	1	RRPO_EAV	P19811	equine arte
43	177.5	2.0	2492	1	POLN_EEAVT	P27282	venezuelan
44	176	2.0	1385	1	RRPO_PIAMV	007518	plantago as
45	174.5	1.9	2492	1	POLN_EEVVP	P36328	venezuelan

ALIGNMENTS

RESULT	1	STANDARD	PRT	1693 AA.
ID	POLN_HEVPA			
AC	P33424;			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	15-DEC-1998 (Rel. 37, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	NON-STRUCTURAL POLYPEPTIDE [CONTAINS: RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48); HELICASE].			
OS	Hepatitis E virus (strain Pakistan) (HEV).			
OC	Viruses; ssRNA positive-strand viruses, no DNA stage.			
OX	NCBI_TaxID=33774;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92115700; PubMed=1731327;			
RA	Tsarev S.A., Emerson S.U., Reyes G.R., Tsareva T.S., Legters L.J.,			
RA	Malik I.A., Iqbal M., Purcell R.H.;			
RT	"Characterization of a prototype strain of hepatitis E virus.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 89:559-563(1992).			
CC	-I- MISCELLANEOUS: HEPATITIS E VIRUS IS THE MAJOR CAUSATIVE AGENT OF ENTERICALLY TRANSMITTED NON-A, NON-B HEPATITIS (ET-NANBH).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: M80581; AAA45725.1; -			
DR	MEROPS: C41.001; -			
DR	InterPro: IPR002589; DUF27.			
DR	InterPro: IPR002588; V_methyltransf.			
DR	InterPro: IPR000606; Viral_helisel.			
DR	Pfam: PF01661; DUF27; 1.			
DR	Pfam: PF01443; Viral_helisel; 1.			
DR	Pfam: PF01660; Vmethyltransf; 1.			
DR	SMART: SM00506; Atpp; 1.			
KW	Polyprotein; Transferase; RNA-directed RNA polymerase; Helicase;			
KW	ATP-binding.			
FT	NE-BIND 975 982 ATP (POTENTIAL).			
SO	SEQUENCE 1693 AA; 185149 MW; 5A0F03FB1F199E8 CRC64;			

Query Match 82.8%; Score 7459; DB 1; Length 1693;
Best Local Similarity 81.8%; Pred. No. 0;
Matches 1401; Conservative 109; Mismatches 160; Indels 42; Gaps 7;

QY	1	PGTTTAEQALAAANSALANAVVVFSLRSVQTEILINIMQPRQVLFREVTLMNPQR	60
DB	10	PGTTTAEQALAAANSALANAVVVFSLRSVQTEILINIMQPRQVLFREVTLMNPQR	69
QY	61	VHNELEVCRAAGRCLEAGPRSTINDPNVLRHCFELRPAQVDQVRSATPGRPAAN	120
DB	70	VHNELEVCRAAGRCLEAGPRSTINDPNVLRHCFELRPAQVDQVRSATPGRPAAN	129

```

121 CRRSALRGLPADRTVCEFGFSCAFEAETGVALYSLHDLMPADVAEAMARHCXRLTYAA 180
130 CRRSALRGLPADRTVCEFGFSCAFEAETGVALYSLHDLMSPDAEAMFRRHGMTLYAA 189
181 LHLPEVLLPGTYHTTYSLLIHDGDAVVTYEGDTSAGYNHDSILRAMIRTTKIVGDH 240
190 LHLPEVLLPGTYHTTYSLLIHDGDAVVTYEGDTSAGYNHDSILRAMIRTTKIVGDH 249
241 PLVIEVRAICGHFVLLTLAAPERPMYVYPRSTREYVVRSLFEGGSSPSLFPSCSK 300
250 PLVIEVRAICGHFVLLTLAAPERPMYVYPRSTREYVVRSLFEGGSSPSLFPSCSK 309
301 STEHAPVIMIRMLMFGATLDDQAFCCSRMLTYLIGISYKTVGLVANEENMASEDL 360
310 STEHAPVIMIRMLMFGATLDDQAFCCSRMLTYLIGISYKTVGLVANEENMASEDL 369
361 TAXITAAVLTJCHQRYLTQAISSKGRRLGVEHAQKFTRLYSWLFKSGRDYIPGRLO 420
370 TAXITAAVLTJCHQRYLTQAISSKGRRLGVEHAQKFTRLYSWLFKSGRDYIPGRLO 429
421 FTAQCRRLSAGFHLDPRLVFEDESPPCGRFTFLKVAACKFCCFMWLOQECCLFEPAB 480
430 FTAQCRRLSAGFHLDPRLVFEDESPPCGRFTFLKVAACKFCCFMWLOQECCLFEPAB 489
481 GLVGDGHNEAVEGSEVDPAEPALDVSTYAVHGHOLEALYRALNVPDODIARASRLT 540
490 GLVGDGHNEAVEGSEVDPAEPALDVSTYAVHGHOLEALYRALNVPDODIARASRLT 549
541 ATVEVLASDRLECRVYLNKTFRTTVVDGAHLLEANGPEQYVLSFPAOSKSGASHLT 600
550 ATVEVLASDRLECRVYLNKTFRTTVVDGAHLLEANGPEQYVLSFPAOSKSGASHLT 609
601 YELTPAGLOVRISNGLDCTATPPPGAPSAAPGEVAACSAALYRNRTOHSLTGGWL 660
610 YASASAGLEVRYAAGLDHRAVFAFGVSPRSAPGEVTAFCASALYRNRTOHSLTGGWL 669
661 LHPBGLLGPFPSPGHMESANPFCGEGLYTRTWS-----TSFGSDSPPE 709
670 LHPBGLLGPFPSPGHMESANPFCGEGLYTRTWSVDAVPSAPQDLGFTS-----E 724
710 AAAPMAATPGLPHSTPPYSDIWLPRPSEEROVDAAPV-PRAPDA-GLPGVYLTTPP 767
725 PSLPSRAATP-----TP-----AAAPRPPADPSPLTSAAPARCEAP 761
768 PPPVHAPSTP-PPSRNRRLTYTPDGAKYVAGSLFESDDMLVANSNPGHRCGGCHAF 826
768 GATAPAPALTHTOTARHRLTLTPDGSKYVAGSLFESTCTWLNVASNVDRHREGGLCHAF 821
827 YORFPEAFYTFEIMREGLAAYTLPRPLIHAVADRYEONPKRLEAAVRETCSRGT 886
822 YORFPEAFYTFEIMREGLAAYTLPRPLIHAVADRYEONPKRLEAAVRETCSRGT 881
887 APLUGSGIYQVPSLSEFAMEBNRHPGDELYTERPAANWFEMAKRQAPVLTTEBTANT 946
882 APLUGSGIYQVPSLSEFAMEBNRHPGDELYTERPAANWFEMAKRQAPVLTTEBTANT 941
947 ANLALEIDATEYGRACACCTISPGIVHVOFTAGVPGSKSRISIOGDVYVVPRELR 1006
942 ANLALEIDATEYGRACACCTISPGIVHVOFTAGVPGSKSRISIOGDVYVVPRELR 1001
1007 NSMRRGGAATPHTAARTIGRRVVIDEAPSLPHLLILHMQRASSVHLGDPNQIPAI 1066
1002 NSMRRGGAATPHTAARTIGRRVVIDEAPSLPHLLILHMQRASSVHLGDPNQIPAI 1061
1067 DEHAGLVAPRPELAPTSWMXVTHRCPADVCELLGAVPKIQTTSRVLSLFWNPATG 1126
1062 DEHAGLVAPRPELAPTSWMXVTHRCPADVCELLGAVPKIQTTSRVLSLFWNPATG 1121
1127 OKLVYQAQKANKPGALTYHEAGAFETFTTITATADAGLIQSSRAHAIVALTRETEGC 1186
1122 OKLVYQAQKANKPGALTYHEAGAFETFTTITATADAGLIQSSRAHAIVALTRETEGC 1181

```

```

1187 VIIDAPGLLEVEGIDSPVIVNNEFLAGGEVXHRPSVYIPRGNPDONLGLTOAEPSCQISA 1246
1182 VIIDAPGLLEVEGIDSPVIVNNEFLAGGEVXHRPSVYIPRGNPDONLGLTOAEPSCQISA 1241
1247 YHQLAEELGHRAPAAVAVLPCELEQGLLYMPDELTYSDSVLELTDIVICRMAAPSO 1306
1242 FHQLAEELGHRAPAAVAVLPCELEQGLLYMPDELTYSDSVLELTDIVICRMAAPSO 1301
1307 RRAVISTIVGRYGRTKLEAASDVRESLAFPIPIPGVOATCELYELVAMEVEKGD 1366
1302 RRAVISTIVGRYGRTKLEAASDVRESLAFPIPIPGVOATCELYELVAMEVEKGD 1361
1367 GSAVLELDLCNDRVSKRTPEOKXCKFTTGETTIAHGKVGQGISANSKTECALFGPFRAT 1426
1362 GSAVLELDLCNDRVSKRTPEOKXCKFTTGETTIAHGKVGQGISANSKTECALFGPFRAT 1421
1427 EKEITALLPPIFYGDATYESFPAAVSAGSCMPFENDFSEFDSTONNESLGLCYWHE 1486
1422 EKEITALLPPIFYGDATYESFPAAVSAGSCMPFENDFSEFDSTONNESLGLCYWHE 1481
1487 ECGMPOMLRLYHLVRSAMIILOAPKESLKGPMKKSGBEPTLLMNTVMNALIACYEPFR 1546
1482 ECGMPOMLRLYHLVRSAMIILOAPKESLKGPMKKSGBEPTLLMNTVMNALIACYEPFR 1541
1547 DERVAAFKGDSDSVLCSDYRQSRMAAIIAGGLKLYDYRPIGLYAGVVAAPGLGTLPD 1606
1542 DERVAAFKGDSDSVLCSDYRQSRMAAIIAGGLKLYDYRPIGLYAGVVAAPGLGTLPD 1601
1607 VVRPAGRLSEKMWGQPPREABDLAVCELTGNTVAVOVADVYSRYVSGPGLVHNI 1666
1602 VVRPAGRLSEKMWGQPPREABDLAVCELTGNTVAVOVADVYSRYVSGPGLVHNI 1661
1667 GMLQITADGKAHFTETIKPVLDTJNSIIOERVE 1698
1662 GMLQITADGKAHFTETIKPVLDTJNSIIOERVE 1693

POLN_HEVBU
POLN_HEVBU STANDARD: PRG: 1693 AA.
AC P29324:
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NON-STRUCTURAL POLYPEPTIDE [CONTAINS: RNA-DIRECTED RNA POLYMERASE
(LOC 2.7.7.48); HELICASE].
OS Hepatitis E virus (strain Burma) (HEV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
OX NCBI_TaxID=31767;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92024067; PubMed=1926770;
RA Tam A.W., Smith M.M., Guerra M.E., Huang C.-C., Bradley D.W.,
RT Fry K.E., Reyes G.R.;
RT "Hepatitis E virus (HEV): molecular cloning and sequencing of the
RT full-length viral genome.";
RT Virology 185:120-131(1991).
RC -; MISCELLANEOUS; HEPATITIS E VIRUS IS THE MAJOR CAUSATIVE AGENT OF
CC ENTERICALLY TRANSMITTED NON-A, NON-B HEPATITIS (EH-NANBH).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M73218; AAA5734.1;
DR PIR: A40778; MNMWE.
DR MEROPS: C41.001;
DR InterPro: IPR002589; DUF27.
DR InterPro: IPR002588; V_methyltransf.

```

DR InterPro: IPR000606; Viral_helicase.
DR Pfam: PF01661; DUF27.1.
DR Pfam: PF01443; Viral_helicase1.1.
DR Pfam: PF01660; Vmethyltransf.1.
DR SMART: SM00506; ATP.1.
KW Polyprotein; Transferase; RNA-directed RNA polymerase; Helicase;
KW ATP-binding.
FT NP_BIND 975 982 ATP (POTENTIAL).
SQ SEQUENCE 1693 AA; 185191 MW; 2F355E46E9ED198 CRC64.

Query Match 82.68; Score 7448; DB 1; Length 1693;
Best Local Similarity 81.88; Pred. No. 0;
Matches 1396; Conservative 110; Mismatches 169; Indels 32; Gaps 6;

QY 1 PGTTAIEOALAAANSALANAVVPPPLSRVOTELLINIMOPOLVREPEVLNHP10R 60
DB 10 PGTTAIEOALAAANSALANAVVPPPLSRVOTELLINIMOPOLVREPEVLNHP10R 69
QY 61 VIHNELOYCARAGRCLEVGAPHRSTINDNPNVLRHCELRPVGRDVGWYAPTRGPAN 120
DB 70 VIHNELEYCARAGRCLEVGAPHRSTINDNPNVLRHCELRPVGRDVGWYAPTRGPAN 129
QY 121 CRRSALGLPDPADRTYCFDEFSRCFAETGVALYSLHDLMPADYAEAMARHGXTLYAA 180
DB 130 CRRSALGLPDPADRTYCFDEFSRCFAETGVALYSLHDLMPADYAEAMARHGXTLYAA 189
QY 181 LHLPEVLLPGYHTTSTYLLIHNDGRAVYVEGDTSGAGYNDVSIIRAMIRTKYGDH 240
DB 190 LHLPEVLLPGYHTTSTYLLIHNDGRAVYVEGDTSGAGYNDVSIIRAMIRTKYGDH 249
QY 241 PLYTERRAIGCHVLLITAAPEPSMPYVYPRSTEVYVASTIGPGSGSLFSPASCTK 300
DB 250 PLYTERRAIGCHVLLITAAPEPSMPYVYPRSTEVYVASTIGPGSGSLFSPASCTK 309
QY 301 STEHAVVNIHMDRLMFGATLDDOAFCCSRMTYLRGISYVWYVGAIVANGMNASDAL 360
DB 310 STEHAVVNIHMDRLMFGATLDDOAFCCSRMTYLRGISYVWYVGAIVANGMNASDAL 369
QY 361 TAKITAAVLTGHQRYLRTQAISKGMRLGVEHAKETITRLYSWLFKSGRDYIPGRLO 420
DB 370 TAKITAAVLTGHQRYLRTQAISKGMRLGVEHAKETITRLYSWLFKSGRDYIPGRLO 429
QY 421 FYACCRWLSAGHFLDRVLVFEDESVCRCRTFLKKAAGKFCCKMRLGOCCTFLPAE 480
DB 430 FYACCRWLSAGHFLDRVLVFEDESVCRCRTFLKKAAGKFCCKMRLGOCCTFLPAE 489
QY 481 GLVGDHGHDEAYEGSEVDPAPAHLDVSGTYAVAHGQLEALYRALVNPDIARASRLT 540
DB 490 GAVGDGHGHDEAYEGSEVDPAPAHLDVSGTYAVAHGQLEALYRALVNPDIARASRLT 549
QY 541 ATVELVASPRLCEKRTVLGNKTRFTTVYDGAHLBANGPEOYVLSFDSASROSMGASHLT 600
DB 550 ATVELVASPRLCEKRTVLGNKTRFTTVYDGAHLBANGPEOYVLSFDSASROSMGASHLT 609
QY 601 YELTPAGLOVRISNGDLCRTATPPGAPSAPEVAAFGSCALYRYNRFOTRHSLTGLW 660
DB 610 YELTPAGLOVRISNGDLCRTATPPGAPSAPEVAAFGSCALYRYNRFOTRHSLTGLW 669
QY 661 LHPGLIGTFPPSPGHIMESANPFCGEGTLTYRTWS-TSGFSDFSP-----PEAAPA 714
DB 670 FHPBGLIGLTFPPSPGHIMESANPFCGEGTLTYRTWS-EVDAVSSPARDLGFMSEPSLPS 729
QY 715 MAATPGLPSTPVSIDIMVLRPPSEEFQVDAAPV-PPAPPPAGLP-GPVVLTTPPPPPVH 772
DB 730 RAATPPLT-----AAPLPAPDPSPPSPAPALAEPAASGATAG 766
QY 773 KPSPFP-PPSNRRLLTYTPGCAKYASLSFESDCDMLYNASNPCHRGCGGICAHFYORP 831
DB 767 APATHTGTARRRLFTFYPPGSKVFAGSLFESTCTLWLNASNDVHRGCGGICAHFYORP 826
QY 832 EAFYPTFEIMREGIAATYTLPRPIIHAVADPYRVEONPKRLEAAYRETCSRGAAPPL 891
DB 832 EAFYPTFEIMREGIAATYTLPRPIIHAVADPYRVEONPKRLEAAYRETCSRGAAPPL 891

DB 827 ASDPAEFVMDGAAATYTLPRPIIHAVADPYRLEHNPKRLEAAYRETCSRGAAPPL 886
QY 892 GSGIYQVPVSLSPAMERNRHPGDELITLPEANMPFANKPAQVLTITETATAMAL 951
DB 887 GTGIYQVPVSGSPFAMERNRHPGDELITLPEANMPFANKPAQVLTITETATAMAL 946
QY 952 EIDATFVGRACACTISPGIVHGYTAGVGGSKSSIOGDDVYVVPFRELNSMR 1011
DB 947 EIDATFVGRACACTISPGIVHGYTAGVGGSKSSIOGDDVYVVPFRELNSMR 1006
QY 1012 RGFAPFTPHPARVYTIARRVVIDEAPSLPHLLLLHMQRSSVHLLDPPNIPAIDEHA 1071
DB 1007 RGFAPFTPHPARVYTIARRVVIDEAPSLPHLLLLHMQRSSVHLLDPPNIPAIDEHA 1066
QY 1072 GLVPAIRPELAPTSWAVYTHRCPADVCELLRGATPKIOTTSTVRLSFWMPEAIGOLVX 1131
DB 1067 GLVPAIRPELAPTSWAVYTHRCPADVCELLRGATPKIOTTSTVRLSFWMPEAIGOLVX 1126
QY 1132 TOAKKANPGATVHEAGATFETETTLIATPADAGLIQSSRAHAIVALTRTEKCVILDA 1191
DB 1127 TOAKKANPGATVHEAGATFETETTLIATPADAGLIQSSRAHAIVALTRTEKCVILDA 1186
QY 1192 PGLREYGISDVIVNNEFLAGGEVGHRSVYIPRGNPDONLGTQAPPPSCQISAYHQLA 1251
DB 1187 PGLREYGISDVIVNNEFLAGGEVGHRSVYIPRGNPDONLGTQAPPPSCQISAYHQLA 1246
QY 1252 EELGHRPAPVAALVPCPELEQGLIYMPQELTVSDSVLVELTDIVHCRMAAPSGORAVL 1311
DB 1247 EELGHRPAPVAALVPCPELEQGLIYMPQELTVSDSVLVELTDIVHCRMAAPSGORAVL 1306
QY 1312 STLVGRGRKRLYEAAHSDVRESLARIPIIGVQATCTGLYLYEAMNEKGGDSAYL 1371
DB 1307 STLVGRGRKRLYEAAHSDVRESLARIPIIGVQATCTGLYLYEAMNEKGGDSAYL 1366
QY 1372 ELDLCNBDVSRITFFOKXCNFTTGETIAHKGVOGSIAMSKTCAFLGPFRAIEREIL 1431
DB 1367 ELDLCNBDVSRITFFOKXCNFTTGETIAHKGVOGSIAMSKTCAFLGPFRAIEREIL 1426
QY 1432 ALDPNIFYGDAEESVFAAAYSGAGSCOMFENDPSEFSTONNFSIGLECVYNECCMP 1491
DB 1427 ALDPNIFYGDAEESVFAAAYSGAGSCOMFENDPSEFSTONNFSIGLECVYNECCMP 1486
QY 1492 QMLRLYHLVNSAMTILQAPKESLKGFWKHSGEGETLLMNTVMMAITIAHXYERDRVA 1551
DB 1487 QMLRLYHLVNSAMTILQAPKESLKGFWKHSGEGETLLMNTVMMAITIAHXYERDRVA 1546
QY 1552 AFKGDSDVILCSDRQSRNMAALITAGCGIKLVYRPITGLAGVVAAPGLGTLDPVYFA 1611
DB 1547 AFKGDSDVILCSDRQSRNMAALITAGCGIKLVYRPITGLAGVVAAPGLGTLDPVYFA 1606
QY 1612 GRLEKMMWGPGERAEOLRLAVCDLRLNVAOVCVDVYSRVYGVSPGLVHNLIGMLQ 1671
DB 1607 GRLEKMMWGPGERAEOLRLAVCDLRLNVAOVCVDVYSRVYGVSPGLVHNLIGMLQ 1666
QY 1672 IADGKAHFTETIKRVLDTNLSIIORVE 1698
DB 1667 IADGKAHFTETIKRVLDTNLSIIORVE 1693

RESULT 3
POLN_HEYME STANDARD; PRT; 1691 AA.
AC 003495;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NON-STRUCTURAL POLYPEPTIDE [CONTAINS: RNA-DIRECTED RNA POLYMERASE
(EC 2.7.7.48); HELICASE].
OS Hepatitis E virus (strain Mexico) (HEV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
OX NCBI_Taxid=31768;
RN [1]
SEQUENCE FROM N.A.

RX MEDLINE-93079857; PubMed-1448913;
 RA Huang C.C., Nguyen D., Fernandez J., Yun K.Y., Fry K.E.,
 RA Bradley D.W., Tam A.W., Reyes G.R.;
 RT "Molecular cloning and sequencing of the Mexico isolate of hepatitis
 RT E virus (HEV).";
 RL Virology 191:550-558(1992).
 RN
 RP SEQUENCE OF 965-1691 FROM N.A.
 RX MEDLINE-92271462; PubMed-1589964;
 RA Fry K.E., Tam A.W., Smith M.M., Kim J.P., Luk K.C., Young L.M.,
 RA Platak M., Feldman R.A., Yun K.Y., Purdy M.A., McCaustland K.A.,
 RA Bradley D.W., Reyes G.R.;
 RT "Hepatitis E virus (HEV): strain variation in the nonstructural gene
 RT region encoding consensus motifs for an RNA-dependent RNA polymerase
 RT and an ATP/GTP binding site.";
 RL Virus Genes 6:173-185(1992).
 CC -1- MISCELLANEOUS: HEPATITIS E VIRUS IS THE MAJOR CAUSATIVE AGENT OF
 CC ENTERICALLY TRANSMITTED NON-A, NON-B HEPATITIS (ET-NANBH).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL M74506; AAA45730.1; -
 DR PIR: A44212; A44212.
 DR MEROPS: C41.001; -
 DR InterPro: IPR002589; DUF27.
 DR InterPro: IPR002588; V_methyltransf.
 DR InterPro: IPR000606; Viral_helices1.
 DR Pfam: PF01661; DUF27.1.
 DR Pfam: PF01443; Viral_helices1.1.
 DR Pfam: PF01660; Vmethyltransf.1.
 DR SMART: SM00506; A1pp.1.
 DR Polyprotein: Transferase; RNA-directed RNA polymerase; Helicase;
 KM ATP-binding.
 FT NP_BIND 973 980 ATP (POTENTIAL).
 FT SEQUENCE 1691 AA; 185224 MW; DB3F0B2C91F871B CRC64;
 SQ
 Query Match 82.5%; Score 7433; DB 1; Length 1691;
 Best Local Similarity 81.0%; Pred. No. 0;
 Matches 1384; Conservative 121; Mismatches 166; Indels 38; Gaps 5;

Db 370 TAVITAAVLTICHORYLRTQALSKGRRLLEHQAQFISRLYSWLFESKGRDYIPGRQLO 429
 QY 421 FTAQCRWMLASGFLHDPRLVDESVPCRCRTFLKVKAGPCCFMKMLQOECTCFLEPAE 480
 Db 430 FYAQCRWMLASGFLHDPRLVDESVPCRCRTTIRIRAKPCFPMKMLQOECSCLQPAE 489
 QY 481 GLVGDHNDNEAYEGSEVDPAEPAILDSGYTAHVGHOLEALYRALNPQDIAARASRLT 540
 Db 490 GLAGDGHNDNEAYEGSDVTPAEPATLIDIGSYIVGDSRLQTYOALDLPADIVARARLS 549
 QY 541 ATVELVASPDRLECRVTYGNKTRFTVYDCAHLEANGPROYVLPDASQSGAGSHSLT 600
 Db 550 ATVTYETISRGIDCQTMIGNKTFLETVGAREVLENGPOLMLSEDSQCSMAAGFCLT 609
 QY 601 YELTPAGIDVRISSNGLDCTATFPGASAPAGEVAACSAALYRNRTQHRSLTGLM 660
 Db 610 YAAVDCGLEVHSTAGLESRYVFPNGNATPAEPSETACSAALYRNRSQSGVIGSLM 669
 QY 661 LHPEGLLGIFPPSPGCHWESANPCGEGTLYTRWST-----SGFSSDSPE 709
 Db 670 LHPEGLLGIFPPSPGHEWRSANPCGEGTLYTRWSTITDPLTVGLISG-HLDAAPHS 728
 QY 710 AAAPAAATPGLPHSTPPVSDIWLVPPESEEOVAAPVPAPDPAGLGPVLPPEPP 769
 Db 729 GGPATATGPAGSSDSPDD--PLEVDIGSRPSGAR-PAGNPQGV----- 774
 QY 770 PVHKSIPPPSRNRRLLYTPDCAKVVAGSLFESDCDMLVNASNPGRRGGGLCHAFYOR 829
 Db 775 -----QRLHLHYTPDQAKIVGSIYESECTWLVNASNGHRGGGLCHAFYOR 822
 QY 830 FPEATPTFEFIMREGIATLTLPRLTHAVADYRREONPKRLEAAYRETCSRGRATP 889
 Db 823 YPDSFPAETKFWNRDGLAATLTLPRLTHAVADYRLEHNPKRLEAAYRETCSRGRATP 882
 QY 890 LUGSGIYQVPSLSPDAMRNHRPQDELUTLTPAAMVPAKPAOVLTYTSDARTANL 949
 Db 883 LUGAGIYQVPSLSPDAMRNHRPQDELUTLTPAAMVPAKPAOVLTYTSDARTANL 942
 QY 950 ALEIDATFVGRACAGCTTSPGIVHGYQFAGYVGSKRSISQGDVDDVVVVPTRRLNSM 1009
 Db 943 ALEIDSGSEVGRACAGCKKPCGVVRQYFAGYVGSKRSISQGDVDDVVVVPTRRLNSM 1002
 QY 1010 RRRGFATPTPHTAARTYIGRRVVIDEAPSLPHLLILHMQRASSVHLGDPNOIPAIDE 1069
 Db 1003 RRRGFATPTPHTAARTYIGRRVVIDEAPSLPHLLILHMQRASSVHLGDPNOIPAIDE 1062
 QY 1070 HNGVPAITPELAPTSWNVTHRCPADVCGLRGATPKIOTTSRYLRSLFNNRPAIGQL 1129
 Db 1063 HNGVPAITPELAPTSWNVTHRCPADVCGLRGATPKIOTTSRYLRSLFNNRPAIGQL 1122
 QY 1130 VYTOAAKAAKPAITVHEOGATFTETIATADARGLIOSSRAHAIVALTFRTEKCVTL 1189
 Db 1123 VYTOAAKAAKPAITVHEOGATFTETIATADARGLIOSSRAHAIVALTFRTEKCVTL 1182
 QY 1190 DAPGLIREVIGSDIVYNNFLAGGEVGHHPSPVIRPGNDONLGLQAFPPSCQISAVHQ 1249
 Db 1183 DSPGLIREVIGSDIVYNNFLAGGEVGHHPSPVIRPGNDONLGLQAFPPSCQISAVHQ 1242
 QY 1250 LAEBELGRHAPAAVLPPEPELQGLLYMPQELIYSDSVLVEELDIYVCRMAAASQRA 1309
 Db 1243 LAEBELGRHAPAAVLPPEPELQGLLYMPQELIYSDSVLVEELDIYVCRMAAASQRA 1302
 QY 1310 VLSITLVGRYGRRTKYLEAHSDVRESLAFITPLIGVQATCQELVELYVAMKEGQDGA 1369
 Db 1303 VLSITLVGRYGRRTKYLEAHSDVRESLAFITPLIGVQATCQELVELYVAMKEGQDGA 1362
 QY 1370 VLELDICNRDVSRTFFQCKNKFETGETIAGKVGQGISANSKFCALFGMPRAIEKE 1429
 Db 1363 VLELDICNRDVSRTFFQCKNKFETGETIAGKVGQGISANSKFCALFGMPRAIEKE 1422
 QY 1430 ITALLPNTFYGDATVEESFAAASVAGAGSCMYFENDSFSDTONNLSIGLECYAMEECG 1489
 Db 1423 ITSLPQAVFYGDATVDSFAAASVAGAGSHAWFENDSFSDTONNLSIGLECYAMEECG 1482

QY 1490 MPQWLRVHLVRSAMTLLQAPKESLGFKKHSGERGCTLLMNTVMNMATIAHGEFRDR 1549
DB 1483 MPQWLRVHLVRSAMTLLQAPKESLGFKKHSGERGCTLLMNTVMNMATIAHGEFRDR 1542
QY 1550 VAAFKGDDSVYLCSDRROSNAALTAGGGLKLVYRPIGLTAGVVAAGLTLDVVR 1609
DB 1543 VAAFKGDDSVYLCSEYRQSPGASLTAGGGLKLVYRPIGLTAGVVAAGLTLDVVR 1602
QY 1610 FAGRLSEKMWGPBPERAEQRLAVCDFELRGLTNVAQCVDVYSRVYGVSPGLVHNLIGML 1669
DB 1603 FAGRLSEKMWGPBPERAEQRLAVCDFELRGLTNVAQCVDVYSRVYGVSPGLVHNLIGML 1662
QY 1670 QTIADGKAHFTETIKVPYLDITNSIORVE 1698
DB 1663 QTIADGKAHFTESVKRILDITSHIMARSE 1691

RESULT 4
POLN_HEVMT STANDARD: PRT; 1693 AA.
ID POLN_HEVMT 004610;
AC 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NON-STRUCTURAL POLYPEPTIDE [CONTAINS: RNA-DIRECTED RNA POLYMERASE
(EC 2.7.7.48); HELICASE].
OS Hepatitis E virus (strain Myanmar) (HEV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
OX NCBI_Taxid=31769;
RN [1]
RX SEQUENCE FROM N.A.
RA MEDLINE=9322753; PubMed=8470371;
RA Aye T.T., Uchida T., Ma M.Z., Iida F., Shikata T., Ichikawa M.,
RA Rikihisa T., Wain K.;
RT "Sequence and gene structure of the hepatitis E virus isolated from
RT Myanmar";
RT Virus Genes 7:95-109(1993).
RL -1- MISCELLANEOUS: HEPATITIS E VIRUS IS THE MAJOR CAUSATIVE AGENT OF
CC ENTERICALLY TRANSMITTED NON-A, NON-B HEPATITIS (ET-NANBH).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib.ch).
CC
CC EMBL: D10330; BAA01172.1; -
DR MEROPS: C41.001; -
DR InterPro: IPR002589; DUF27.
DR InterPro: IPR002588; V_methyltransf.
DR InterPro: IPR000606; Viral_helices.
DR Pfam: PF01661; DUF27.1.
DR Pfam: PF01443; Viral_helices.1.
DR Pfam: PF01660; V_methyltransf.1.
DR SMART: SM00506; A1pp; 1.
KW Polyprotein; Transferase; RNA-directed RNA polymerase; Helicase;
KW ATP-binding.
FT NP_BIND 975 982 ATP (POTENTIAL).
SQ SEQUENCE 1693 AA; 185215 MW; AAB4C9140N7E21EA CRC64;

Query Match 82.5%; Score 7433; DB 1; Length 1693;
Best Local Similarity 81.1%; Pred. No. 0;
Matches 1394; Conservative 109; Mismatches 159; Indels 56; Gaps 6;

QY 2 GITTAEQAALAAANSAANAVVVRPLSLRVQTEIINLMQRPQVLRPEVLMNHPRIQRY 61
DB 11 GITTAEQAALAAANSAANAVVVRPLSLRVQTEIINLMQRPQVLRPEVLMNHPRIQRY 70
QY 62 INHELEQYCRARAGRCLEVGAPHSRINDPNVNLHRCFLRPVGRDVGKWSAPTRGPAANC 121

DB 71 INHELELYCRARAGRCLEVGAPHSRINDPNVNLHRCFLRPVGRDVGKWSAPTRGPAANC 130
QY 122 RSRLRGLEPADRYTCDFGFSRCAFAETGVALSYLHDLMPADVAEMARHGRTLYAAL 181
DB 131 RSRLRGLEPADRYTCDFGFSRCAFAETGVALSYLHDLMPADVAEMARHGRTLYAAL 180
QY 182 HLPPEVLLPPEGYHTTSTYLLIHGDRVVYTYEEDTSAGYHNDVSIIRAWTRTKIYGDHP 241
DB 191 HLPPEVLLPPEGYHTTSTYLLIHGDRVVYTYEEDTSAGYHNDVSIIRAWTRTKIYGDHP 250
QY 242 LVIERVRAIGCHVLLTLTAPESPMPYVPRSTEVYVASTIGPGSGSLFSPASCTKS 301
DB 251 LVIERVRAIGCHVLLTLTAPESPMPYVPRSTEVYVASTIGPGSGSLFSPASCTKS 310
QY 302 TFEHAVVHIMDRMLMFGATLDDQAFCCSRMTYLRGISYKVTYGAALVANEGMASEDALT 361
DB 311 TFEHAVVHIMDRMLMFGATLDDQAFCCSRMTYLRGISYKVTYGAALVANEGMASEDALT 370
QY 362 AXITTAAYLTICHOYRTQALISGMRLGVEAOKETITRLYSWLFKSGRDYIPGQLOP 421
DB 371 AXITTAAYLTICHOYRTQALISGMRLGVEAOKETITRLYSWLFKSGRDYIPGQLOP 430
QY 422 YACCRWLSAGFHLDRVLVFEDESVCRCRTFLKVAAGKCCPMRLVGOCTCFLEPARS 481
DB 431 YACCRWLSAGFHLDRVLVFEDESVCRCRTFLKVAAGKCCPMRLVGOCTCFLEPARS 490
QY 482 LVGDHGDNEAVEYGESEVDPAPPAHLDVSGTYAAGHOLEALYALVPODIAARSLTA 541
DB 491 LVGDHGDNEAVEYGESEVDPAPPAHLDVSGTYAAGHOLEALYALVPODIAARSLTA 550
QY 542 TVELVASPDRLSECRIVLGNKTEFTVVDGAAHLEANGBEQYVLSFDSRQSMAGSHSLY 601
DB 551 TVELVASPDRLSECRIVLGNKTEFTVVDGAAHLEANGBEQYVLSFDSRQSMAGSHSLY 610
QY 602 ELTPRAGLOVRISSGDCSTAPPGGAPAPAEVAFCAALRYRFTGRHSLTGLTML 661
DB 611 AASAAGLEVRVYVAAGLDHRAVFAVPSRASPBEVYAFCAALRYRFTGRHSLTGLTML 670
QY 662 HPEGLIGLEFAPSPGHVWESANFCESTLYLTRWSEVDVASPPARDLGMSEPSIPSR 720
DB 671 HPEGLIGLEFAPSPGHVWESANFCESTLYLTRWSEVDVASPPARDLGMSEPSIPSR 730
QY 716 AATPGPLPHSTPPSDIWLVRPPESEFQVDAAPRPPAPDPAGLGPVLTTPPPPPYHKPS 775
DB 731 AATPGPLPHSTPPSDIWLVRPPESEFQVDAAPRPPAPDPAGLGPVLTTPPPPPYHKPS 785
QY 776 IPPP-----SNRRLTYTPDGAKEYAGSLFESDCMLVNASNPGHRPGG 820
DB 785 IPPP-----SNRRLTYTPDGAKEYAGSLFESDCMLVNASNPGHRPGG 835
QY 821 GLCHAEVYQREPEAFYPTETIREGLAAYTLTPRPIIHAAVADYRVQNPRLKLEAAVRETC 880
DB 831 GLCHAEVYQREPEAFYPTETIREGLAAYTLTPRPIIHAAVADYRVQNPRLKLEAAVRETC 885
QY 816 GLCHAEVYQREPEAFYPTETIREGLAAYTLTPRPIIHAAVADYRVQNPRLKLEAAVRETC 875
DB 875 GLCHAEVYQREPEAFYPTETIREGLAAYTLTPRPIIHAAVADYRVQNPRLKLEAAVRETC 880
QY 881 SRGTAAYPLLSGCIYQVPVSLFSDAMERNHNRGDELYLTERPANNFEAKKAPQVLTIT 940
DB 891 SRGTAAYPLLSGCIYQVPVSLFSDAMERNHNRGDELYLTERPANNFEAKKAPQVLTIT 945
QY 941 EDTPARTANLLEIDATEVEVRACAGCTISPGIYHGYOTAGVPSSGSRSTTQGDVAVVY 1000
DB 951 EDTPARTANLLEIDATEVEVRACAGCTISPGIYHGYOTAGVPSSGSRSTTQGDVAVVY 1005
QY 936 EDVARTANLLEIDATEVEVRACAGCTISPGIYHGYOTAGVPSSGSRSTTQGDVAVVY 995
DB 995 EDVARTANLLEIDATEVEVRACAGCTISPGIYHGYOTAGVPSSGSRSTTQGDVAVVY 1000
QY 1001 PTERLNSWRRRGFAAFPTPTAARVTIGRRVVIDEAPSLPPLLHLMORASSVHLGDP 1060
DB 1011 PTERLNSWRRRGFAAFPTPTAARVTIGRRVVIDEAPSLPPLLHLMORASSVHLGDP 1065
QY 996 PTERLNSWRRRGFAAFPTPTAARVTIGRRVVIDEAPSLPPLLHLMORASSVHLGDP 1055
DB 1055 PTERLNSWRRRGFAAFPTPTAARVTIGRRVVIDEAPSLPPLLHLMORASSVHLGDP 1060
QY 1061 NQIPALDEFHAGVYPAIRBELAPTSMWXYTHRCRADVCELIKRAYIKQTSVYLSLFW 1120
DB 1071 NQIPALDEFHAGVYPAIRBELAPTSMWXYTHRCRADVCELIKRAYIKQTSVYLSLFW 1125
QY 1056 NQIPALDEFHAGVYPAIRBELAPTSMWXYTHRCRADVCELIKRAYIKQTSVYLSLFW 1115
DB 1115 NQIPALDEFHAGVYPAIRBELAPTSMWXYTHRCRADVCELIKRAYIKQTSVYLSLFW 1120
QY 1121 NEPAIGOKLVYXDOAKAANGALITVEADGATETTTIATADARGIIOSSRAHAIVALT 1180
DB 1180 NEPAIGOKLVYXDOAKAANGALITVEADGATETTTIATADARGIIOSSRAHAIVALT 1185

Db 1116 GEPANGOKLFTQAKANPGSVTVHEAGATYTTETTTATADARGLIOSSRAHAIVALT 1175
 QY 1181 RHEKCVILADAPGLAREVISPVIYNNFPLAGGEVGHRRPVIPRPNPONTICLQAFP 1240
 Db 1176 RHREKCVIILADAPGLAREVISPVIYNNFPLAGGEVGHRRPVIPRPNPONTICLQAFP 1235
 QY 1241 SCQISAYHOLAEELGHRPAVAALPPCELEGGLLYMPQELTVSDSVLFELTDIVCR 1300
 Db 1236 SCQISAFHOLAEELGHRPAVAALPPCELEGGLLYLPQGLIACDSVTFELTDIVCR 1295
 QY 1301 MAAPSRKAVLSTVGRYRRTKLYNAASHDVDSLAFIPALGPQVTTCELYELVEM 1360
 Db 1296 MAAPNRKAVLSTVGRYRRTKLYNAASHDVDSLAFIPALGPQVTTCELYELVEM 1355
 QY 1361 VEGGOGSAVLELDLCNRPVSRITTFQKCKNFTTGETTAHGKVGOGISAMSKTFCALG 1420
 Db 1356 VEGGOGSAVLELDLCNRPVSRITTFQKCKNFTTGETTAHGKVGOGISAMSKTFCALG 1415
 QY 1421 PWFRAIEKITALLPPIFYGDAYEESVFAAASGAGSCWFEENDESEFSTQNNFSLGL 1480
 Db 1416 PWFRAIEKITALLPQGVFYGDADFDYVSAVAAMKAMFENDESEFSTQNNFSLGL 1475
 QY 1481 ECVYMECCMPQMLIRLYHLINSAMWLLQAPKESLKGFMKKHSGEPGTLIMTYMNAVIT 1540
 Db 1476 ECAIMECCMPQMLIRLYHLINSAMWLLQAPKESLKGFMKKHSGEPGTLIMTYMNAVIT 1535
 QY 1541 HCYEEDFPAFAFKGDSVYLCSDYRQSRNAALIGCGLTKVDYRPIGLTAYGVVAPG 1600
 Db 1536 HCYEDFPAFAFKGDSVYLCSDYRQSRNAALIGCGLTKVDYRPIGLTAYGVVAPG 1595
 QY 1601 LGTLPDVVFAAGLSEKNGPGERAEQLRLAVCDELITNVAQVCDVSVSVYGVSPG 1660
 Db 1596 LGALPDVVFAAGLSEKNGPGERAEQLRLAVCDELITNVAQVCDVSVSVYGVSPG 1655
 QY 1661 LVHNLIGMLQITADGKAHFTETIKPVLDTNSTIQRVE 1698
 Db 1656 LVHNLIGMLQAVADGKAHFTESYKPVLDLTNSTICRVE 1693
 RESULT 5
 ID RPO_CGMVS STANDARD; PRT; 1648 AA.
 AC P19523; Q83208; P89877; P90356;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (186 KDA PROTEIN) [CONTAINS:
 DE METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (129 KDA PROTEIN)].
 OS Cucumbers green mottle mosaic virus (watermelon strain SH) (CGMV), and
 OS Cucumbers green mottle mosaic virus (watermelon strain W) (CGMW).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamoviruses.
 NX NCBI_TaxID=12236, 12237;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SH;
 RX MEDLINE=913114400; PubMed=1856687;
 RA Ugaki M., Tomiyama M., Kakutani T., Hidaka S., Kiguchi T., Nagata R.,
 RA Sato T., Motoyoshi F., Nishiguchi M.;
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1496-1648 FROM N.A.
 RC STRAIN-W;
 RX MEDLINE=89073773; PubMed=3201760;
 RA Saito T., Imai Y., Meshi T., Okada Y.;
 RT "Interviral homologues of the 30k proteins of tobamoviruses."

RL Virology 167:653-656(1988).
 CC - FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
 CC RNA REPLICATION.
 CC - FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
 CC CAPPING AND AN RNA HELICASE.
 CC - MISCELLANEOUS: READTHROUGH OF THE TERMINATOR UAG OCCURS BETWEEN
 CC CODONS FOR LYS-1144 AND GLN-1146.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D12505; BAA18895.1; -
 DR EMBL: D12505; BAA18896.1; -
 DR EMBL: J04322; AAA46382.1; -
 DR PIR: J01157; WMTMS2.
 DR InterPro: IPR001788; RNA_dep_RNAPol2.
 DR InterPro: IPR002588; V_methyltransf.
 DR InterPro: IPR006066; Viral_helicase.
 DR Pfam: PF00978; RNA_dep_RNAPol2; 1.
 DR Pfam: PF01443; Viral_helicase1; 1.
 DR Pfam: PF01660; Vmethyltransf; 1.
 KW Transferrase; RNA-directed RNA polymerase; Helicase; ATP-binding.
 FT CHAIN 1 1648
 FT NP_BIND 863 870
 FT FT 1 144
 FT SEQUENCE 1648 AA; 186549 MW; 1D1AFEEB/65595 CRC64;
 SQ
 Query Match 4.5%; Score 406.5; DB 1; Length 1648;
 Best Local Similarity 18.8%; Pred. No. 4.3e-18;
 Matches 350; Conservative 220; Mismatches 633; Indels 661; Gaps 73;
 QY 77 CLEVA-HPRSINDNPVNHRC-----FLRPGRDVRWYSAPTRGPANCRSALRGLP- 130
 Db 105 CYDIGNTQYHLEFKGRSYHCCNPCLDLKDVARNW-YNDMTITQYHQRKSGGCRPLPT 163
 QY 131 ---PADRY-----CFDGSRCFAFAETG---VALYSLHDLMPADVAEMARRGX 174
 Db 164 FQIDFRRYDSSPCAVTSCSDVFOECSYDPSGRDNHNAVLSHSIDYIPYSSIGPALHKNV 223
 QY 175 TRLYALHLRPVLL--PRGYHTSYLLIHGDRAVYTYBEDTSAGYHDSILRAWR 232
 Db 224 RVCYAFHFSEALLGSPVGNLNSIGAQFRVGDVHFLFSESTLHTSHLENIKLIYM 283
 QY 233 TTKIVGDHPLVIER--VRAIGCHVLLLTAAPEPSMPYVYPRSTEVYVRSIFGPGS 289
 Db 284 RTYFPADRDYVYIKFEMVRVDTFFFRVLVRADTHLHKSQVGHYSKSKSEYFL-----NT 338
 QY 290 PSLPSPACSTKSTF-----HAYVHITM 311
 Db 339 PPIF---ODKATFSVWEPEAKRKVLIPKFEISRFELSGNVKISRLMDADFVHTIINH- 393
 QY 312 DKLMLFGATLDQAFCCSLMTYLLNGISYKYVVGALVNAEGNASEDLATATTAAYLI 371
 Db 394 -----STYDNKALVWKVNSQSFESIRSRVTVNGSVKSENNVAVDQTLDISIFFLIV 446
 QY 372 CHQRYLTQAIKSGMRRLGVEHAQKFTIRLYSMLFEKSGRDYIPGRQLQFYAQCRRMLSA 431
 Db 447 -----KYRKQVIE----- 454
 QY 432 GFHLDPRLVFEDEYPCRCRFLKRVAGKPCCFMRLGOECTCELEPAGLVGDGHONE 491
 Db 455 -----LMSDKVYIEARGLLRFPADSLKSAVGLG-DCV-----VDALVQTGWFPDS 499
 QY 492 AYESGEVDAEPAAHLDVSTYAVHGHOLEALRYALNPQDIARASRLTAYELVASPDR 551
 Db 500 SDELVLVLP-EP-----FWTFSDYLEGMYEA-----DAKIERESVS-ELLASGSD 542

Best Local Similarity 19.2%; Pred. No. 7.5e-16;
Matches 454; Conservative 201; Mismatches 714; Indels 995; Gaps 104

QY	3	ITTAEOALAAANSALANAVVUPRPSRTOELLINIMORQVUPREVLYMNPORVI	62
Db	40	VTTAAOKAIV-----AIVPRVFTOMVS-----DHP--AL	69
QY	63	HNEEOYCARAGRCLEVGNHRS-----INDNPNVLHCFGLREPGVDORWYS-----	111
Db	70	H-AISRYTRR--HWIEWG--PKAELNLLDPSPGL-----LREAVAYERVMALCHRT	118
QY	112	-----APRGRAN-----CRSALRGLRP-----ADRTYC-EDGF	141
Db	119	ARKATATALEAASEAMHADV-----ALRGASGPFYUHNEDVPHGRAVADRCLLYTPM	175
QY	142	SRCAFAETGYALSLNIDMPADVA-----E	167
Db	176	OMCELMRIIDATLLVANDLWVVALAANYGDWMDLGIMNIDHGGCRACORGAGAPTR	235
QY	168	AMARHGTRLYAALHPRREVLLPRGTHTTSYLLIHGDRAVUTYEGDTSAGIHNHVSIL	227
Db	236	GUTRCSTRIYOVL-----PDTHNPRLYXRCGRPLTRCCAAVLESMWAOHCNGH-----	286
QY	228	RAMITTKIVGDHPLVIERVRAIGCHEVLLTLTAAPESPMPYVPRYSTEVYVNSIFGRC	287
Db	287	-----ARNAVACTLPIRNVRLQGS-----ARV-----	311
QY	288	GSPSLFSPACSTKSTENAVRVNIDRLMLLEGATLDDOAFCCSRMLTYLRGISYKVTGAL	347
Db	312	-LPDLVLAEGRRKMSLSLPRVQORMLSTYCKTSLPRAYUSERFKFMALCHSITLAGN	370
QY	348	VANEGMN-----ASEDALFXHTAAVLTICHQVLETPQVIAISKMRPLGEHAKOFTRLYS-	403
Db	371	VLOEGMKCTCEBDALCA-----YVAFKMQSARLAGIMKAKCAADSLV	417
QY	404	-WLEKSGRDYIPGRLOLGYAOCRWLISA-----GHFLDRVLVFDSEVPCRCRTF	453
Db	418	AGWL-----DTI-----WDALKRFGSVPLAREMBEWEODAAVAAFD-----	454
QY	454	LKKVAGKCCSMRYLGGOCSTGLPRABELVGDHNDHMEVAGSVVDAPRAHNDVSTYA	513
Db	455	-----KQLEDEGGRIHLDTVQ-----PKSPREPIATWT	484
QY	514	VHGHOLE-----ALYRALNVPD-----IARAS	537
Db	485	VHAASEDNHCACARCDVPRERPARAGORDDALLIPWLEAERRLRCREMDPEALRAR	544
QY	538	RUTAVELVASRDLEECTVL-----CNKTFRTYVDGAHLEANGDEOY	581
Db	545	ADTAAAPRPPRRPARRYTVLYKRRANHHGPMVLDERGEADALVULCDLQGLRGERH	604
QY	582	VLSFDASKOSGASHSLTYELTLAGLO--VRISSNGLDCTATPRPGASAAERGEVAAF	639
Db	605	-----FAAGAHMCA--QARGLOAFYKV-----VPRPERMADGG--ARA	639
QY	640	CSALYRYNRFRQRIHSLINGMLNHEGILG--IFRRPSGHIWESANPRCSGEBTLTRT	655
Db	640	WAKEFGCAMAOR-----LLGERVAMNHLRYTDDVPOQL-----ALALRT	679
QY	696	WSTGFSSEDFE-----PREAARPAMAAT-----FGLP-----HS	724
Db	680	LAOGGALALSLVRDLPGGAATDAAVTAANYACGRQSAASPRRGDPRRRARRSORHS	739
QY	725	-----TPPVSDIWLVRPSEEFQVDAARVAPRADAGLRGVULTRPPRRPVNK-----	773
Db	740	DARGTTPPARARDPPR-----APSPRAPRAGDVP-----PIAGADRARDAELE	787
QY	774	-----PSIPPRRNR-----RLTYTPDGAKUYAASLESFESCDWL	808
Db	788	VACPSGCPPTSTRADPDSDIYESARAAAGRNHLKROIIMDRPCKV-----V	835
QY	809	VNASNPGRRPGGICHAIFYORFPEAFYPTTEIMREGDAA--YTLTPRP-----	854

Dh	386	VMAANEGLIAGSSGCAIFANATAA-----	LAANCRLAPCPTEGEAVATPGHC	884
Qy	855	----ITHAVP-----DYREONPKLEAYRE-----	TCSRGTAAIPILASGTY--Qy	898
Dh	885	GUTHIHIAVAPRRRPPRALIEEGEALLERAYRIVALAARMAACVACPELLGAGYNSA	944	
Qy	899	PVSIUSPAMERNHPCGELY-----	TEPANMFE	928
Dh	945	AESIRALAAITREPERVRSVSLHCHDRATILTHASVYVAGLAARVBPPTIELAS-CP	1003	
Qy	929	ANKDAQEVLTITTEDTAFTANLALEIDATEVGRACGCTIS-----	969	
Dh	1004	AGDEGRAPORSASPATPLG-----DATAPERGCGCELCRTYTVNDRAVYNLMTERD	1058	
Qy	970	----PGIV-----HYQF-----	977	
Dh	1059	RGATSWAMRIPEVYVYGREHLATHPEPLNITSVLKPAEVRPPRGMCSCDMRCGNHGMQ	1118	
Qy	978	-----TAGVSGSGSRSIQCGDVVVVYPTRELN-SMRRGFAFTPHAA	1023	
Dh	1119	VRCRPSNAHALCRGTGVPRASTRG---GELDPNCTWLAANVAQAACAGAYTSACP	1175	
Qy	1024	RVTIGR-----	RYVIDAP	1037
Dh	1176	KCAGRALISEARHEDFEALSORMSASHADASPDGTGDLPLMETVGCACSRVWVGS	1235	
Qy	1038	SLPHLLILHMORA-----SSVHL-----GPNQI-----PAI-----	1066	
Dh	1236	EAPPDHLVLSLRPNPNQVLEVVARBEGGNPTFCVANGGGRPRYSDPHMLAVP	1295	
Qy	1067	-----DFE-----HAGLVPAIRELAPTS--WMXYT--	1090	
Dh	1296	LSRGGTCATDEGLAQAYVYDLEVRRLGDAMARALASVORPKGPYNIRVWMMAGA	1355	
Qy	1091	-----HRCRA-----DYCELIR-----	GAYPKI--	1108
Dh	1356	GKTRILIAATREDLYCCPNALLHEIOAKLRARDIDIKNAATYERRLTKPLAATRYI	1415	
Qy	1109	-----OTTSRVL-----	BSLF-WNEP-	1123
Dh	1416	DEAFTLGEYCAFAVSQTTAEVICVDROCGPHYANNCRTPVPRMPTERSRHTMRPD	1475	
Qy	1124	-----AIGOKLYXOAKKANPGALVTHE	1147	
Dh	1476	CMAARLAGLDYDIEBERTGFACNLMDRGROYDLHLAFSRETIV--RLHNEGIRATYIRE	1533	
Qy	1148	AQAGTFETTYITATDARGI-IOSSRANAIVALTRHTEKCVLLD-APGLREVGISDVI	1205	
Dh	1354	AQGSVSTACIHNGROSTDVALALTDLIAVSLTRASDALYHLEDEDSLRAGS----	1589	
Qy	1206	NNFLAGGEV-----GXHR-----PSYIPRGNPONIGTIOAPPPSCQISATYQOL	1250	
Dh	1590	--AFLDGALAELEKEVPAGIDRVAAVEAOAPPLP---PADGIREADOVPPCPRLIEELV	1644	
Qy	1251	AEEIGHPPAPVAANVLPCCPELEGGILYMPQELTVSPSYVFEJLTJIVICRYMAAPSORAV	1310	
Dh	1645	FGRAGH---PHYADLNKRYTEGEREVYX--RISHLLNKHNTHEM-----POTER-V	1689	
Qy	1311	LSTLVGRYGRRTKLYEAH--SDVRESLARFP---TIGVOAAT-----CELEYL	1356	
Dh	1690	LSAIVCA-----VRRYRAGEDGSTLFTAVAROHPRPFOIRPPRVTAGVAQOEMWTYLER	1744	
Qy	1357	VEANVEKGOGSNAVLE-----DLGNRVSRITTEFOKXCKMF-----T	1394	
Dh	1745	IDLTDTVYQMGVARELTDTRYARRPELIPAGCMCTAOSLSVPAFLATLKCYDALGCPD	1804	
Qy	1395	TGETIAGKYGOGISAKSFICALGFPPWFALEKELALPPNITYGDAYESVFAAAS	1454	
Dh	1805	EDCIAAGKAGGLEIRAMAKRMVQVMSPHRRAIOKIIMRLARQOFLVAAGHTEPEVDAMQ	1864	
Qy	1455	GAGSCWTFENDESPDSTONNS-----LGLECVMEBCGMQMLIRLYHLVRS	1503	
Dh	1865	AHYTNAIEVDPEFMDNQIARLDVELISALGLPCA--ED-----YRALRA	1912	

QY	776	IPPSRRRLLYUYPOAKVYASILESPDCMVLVNA--SNPHRPGG-----GLCH--	824
Db	776	LPOFTLN-CLISAVSDOTKYSHEEHLWESIQTLLPDSQUSNEETNLGLSTEHLTAHLHY	834
QY	825	-----AFYQRPPEAFYPTFERIMEGELAAYTLTPRPIAHVADRYVEQNPKLEAAYRET	879
Db	835	NFOATVYSDKGPILIFGSDPIKR-----IDITHTGPSPHSFG-KRLGSGP--	881
QY	880	CSRRGTAAYVLLG-----SGIYQVPSLSPDAMERNHRPDELY---LPEAANMFEA	929
Db	882	-SAGHPSDDLIRAMKSFKYSQVNY-LPFS-----EAHNHPISIAKMLISNMKNKPGD	933
QY	930	NKPAQPLVLTIEDOTAKTANALEIDATAYGACAGCTISPG-----IYHYQTAGVPSSG	985
Db	934	VLSLDDVTSQRTGPPRPKERIIOIDHYLDT-----NPKRTPPVHF---AGFACG	981
QY	986	KSRSDIQ-----GDVDVYVPTRELNSMR---RGFAA--FTPTAHAVTTGRGVV	10322
Db	982	KTYITQDLKTKLEKDFR-VSCPTTELRTKMTAMBLHSQSWMRENTWESSILKSRILY	10400
QY	1033	IDEAPSLPHILLHLMORASSVHL---LQDP-----NQIPAIDEFHAGLVP	10757
Db	1041	IDETIKMPRGVLDLSILADPALELYIILGDPLOGEYHSQSKDSSNHRPS---ETRLRLP	10972
QY	1076	AIRELAPTSWXYTHNRCPADVCELI-----KRAYKIOT-----TSRLASLE	11199
Db	1098	YI-----DMYCMW--SYRIPCIARLEQIHSFNAMQCVISVSTPHDSQVPLNSHASSLT	11515
QY	1120	WNEPAIGOKLVYQAKKAANRGATITHEAGAFETTTITANADARGLQSSRAHAYAL	11797
Db	1152	FN--SLGYR-----SCTISSSGLFPCCPAITIVLDNTKWLSS--ANGVAL	11944
QY	1180	TR-----HTEKCVLLA--PGLLREVGISDIYVNNFEL	12110
Db	1195	TRNSGVQVQFMGPSSYVGCTGSSAMSDAFNNSLIMDYFPFSLPQL---KLITSPILT	12511
QY	1211	AGEVGVXHRPS-----VIRGNPDONIGTLOAPPS	12411
Db	1252	RGPLINATGPSSASTHRSFNFHLRPHILPSYDRDEVTWNPTRPDGPETRLDT--HFLRP	13099
QY	1242	QOIAHYHQALAEELGHRAPAAVAVLPCCPELQGLAMQELVNSQSVLVEL-----T	12942
Db	1310	SRLLPHFDLPRAI--TPRPVSTVDP-POAKASPVY-PGEF--FDSLAAFPLPAHDPSR	13633
QY	1295	DIVACRMAAPS---QKAVLSTLVGRYGRRTKLYEAAH-----SDVRESIARFI	13400
Db	1364	EILKKDOSSNQFMPEDRPFSLSC-----QPSLSIAKNAHRNHDPTLLPASINKKL-RFR	14161
QY	1341	PTIGPVQATP-----C-----ELYELVAMERYGQDSAVILE	13737
Db	1417	PSDSPHOITADVDVILQLEFHSICRAVYSRQPNSTVPFNPETLAEICISLNEYAQLSSTQS	14767
QY	1373	LDCNROYV-----RITFFOKXCKFTTGETIANGKYGGQGISAMSKFFCALPSP---W	14222
Db	1477	TIVNARSRSDPDRHTTKYIFAAQNRKVNDSIFEGSWKACQTLALMHDIYVILVLPVKY	15366
QY	1423	FRAIEKEIILLRPNIP-----YGAAYEESVAAAVSAGASQWYFENDESEFSDTO	14723
Db	1537	QRIDN---ADRPNNIYSHCGKTPNQLRBMQCQHL-----THSPKIANDYTAFDOSQ	15868
QY	1474	NNFSLGLECYVMEECGMPWLIRLY-HLYRSANILOAPKESLKGEMKSHGEPGLLMNT	15333
Db	1587	HGESVIVLEAKMKRLNIPSHILOHLHKLNTVSTOGPLTCM-----RLTGEPCGYDUNT	16411
QY	1533	VMNAAITAHCEPFRDPVAAFKGDDSVYLCSDV--RQSNMAAALLAGCLAKIKVY-----	15868
Db	1642	DYNLAIVISQIDQSCPIWY-SCGDSLIT---DHRLPTRRDWPVSYLKRLHKLREKLETSHP	16971
QY	1587	-----RPIGLAGVAVYAGLGTLPD-----VVRP--AGRLSEKNMGCPBER	16252
Db	1698	LFGCYVYVGRAGCINPLATLCKMLIAVNDADLDRRLSTVLTETTGHLGLESYMLHLPET	17515
QY	1626	AEQURLAVCDF	1636

	RESULT	9	
AC	POLR_TYMC	STANDARD:	PRT: 1844 AA.
Dt	01-DEC-1992 (Rel. 24, Created)		
Dt	01-DEC-1992 (Rel. 24, Last sequence update)		
Dt	15-DEC-1998 (Rel. 37, Last annotation update)		
DE	RNA REPLICASE POLYPROTEIN (EC 2.7.7.48).		
OS	Turnip yellow mosaic virus (isolate TYMC).		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Tymovirus.		
OX	NCBI TaxID=31751;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=92119261; PubMed=1731998;		
RA	Dreher T.W., Bransom K.L.;		
RT	"Genomic RNA sequence of turnip yellow mosaic virus isolate TYMC, a		
RL	cDNA-based clone which verified infectivity.";		
PL	Plant Mol. Biol. 18:403-406(1992).		
-I-	MISCELLANEOUS: THE 206 KDA PROTEIN IS POTENTIALLY A POLYPROTEIN		
CC	(BY HOMOMOLOGY TO THE LONGER PROTEIN OF THE SINDBIS VIRUS).		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its use by		
CC	non-profit institutions as long as its content is in no way used for commercial		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see http://www.isb-sdb.ch/announce/		
CC	or send an email to license@sib-sdb.ch).		
CC	-----		
DR	EMBL: X16378; CAA34415.1; .		
DR	PIR: S19152; S19152.		
DR	InterPro: IPR000606; Viral_helices.		
DR	Pfam: PF01443; Viral_helices; 1.		
KM	Transferrase; RNA-directed RNA polymerase; Polypeptin; ATP-binding.		
FT	NP_BIND 976 983 ATP (BY SIMILARITY).		
FT	DOMAIN 1631 1665 POLYMERASE SITE (BY SIMILARITY).		
FO	SEQUENCE 1844 AA; 206612 MW; 02CB928FCFCCA5EAL CRC64;		
Query Match	3.8%; Score 346; DB 1; Length 1844;		
Best Local Similarity	20.5%; Pred. NO. 4e-14;		
Matches 380; Conservative 210; Mismatches 660; Indels 608; Gaps	94,		
OY	157 LMD---LMPADVAEMAMH-CETRIYAALHPPEVLPPG-----TYHTTSYLIND 204		
Db	::::: : : : ::::: :		
OY	141 MHDAIMTYHPSOIMDLFLKPNMLERIYASLVVPEAHLSDOSFPKLYTTTTRHTLHY- 199		
OY	205 GBRVVATVEGDVSAGYNHDVSYLLRAMIRPTKI-VGNPHVIEVRKAIG-CHEVILLTAAP 262		
Db	200 -----VEPGHGAGSINOP-SDAHSMURLNSTRLGNNHLSTVILLESWGYSILLIORGRP 252		
OY	263 EPSPMYPVP-----YPRST-----EYVNSIFG 285		
Db	253 PDPSPSIQAPPTLMASDLEFRSYOEPRLDVYSFRIPDAIELPQAFTGLQPLRDRLRVPAVVN 312		
OY	286 P-----GGSPSLFPSACSFKSTFAHVPNIWPRMLFGATLDDQAFCCSRIMT 333		
Db	313 ALFTYTRAVRTLRTSDPAFFVMHSKRPDHWDVTSNAMDNLTQE-----ALLNV 361		
OY	334 YLR-GISYKVTVG-----ALVANEGNASSEDALTAXITAYATLICHQRYLTRQAIKSGR 387		
Db	362 PLRPNVYHVIVLOSPIASLSLYLRQHRRKL-TATAVPLISFLTLL-QRL----- 408		
OY	388 RLGVENHAQ-KETTRLYSWLFEKSGRDYIPRGOLFYAQCRRMLSAGFNLDPRVLVFDES 446		
Db	409 PLRPLAEVKYSITAFRRBELYRKKE-----PHNPLDY-----FHQHRIRINYSAI 453		
OY	447 PCRCRFTLKVGAKGCCCFMRNLGOECTCTFLAEABGLVDGHGNDNEAYESSEVDPAEPAL 506		

Db 454 -----SAMPSPHQ 464
 QY 507 DVSQYAVHGHOLEALYRALNVPDIAARA-----SRLATYELVASPDR--- 551
 Db 465 KLP-----HALQKALLLRLPILATPFERSEQSMJPNALSWLTKRFLPQMSL 519
 QY 552-----LECRVYLGKTRRTTYVDGAHLEANGPEQYVLSFDSRSGMAGSHLSLEYLTPAG 607
 Db 520 VLLALESSSTLLHKLFSPPLOAOH-----DYYHRLHFGSYSLQWERP-- 564
 QY 608 LOYRISNGCL---DCTATPPGAPSAAPGEVAFCALRYNRFOR-----HSLGGLW 660
 Db 565 LSTPRTATLPRTPTSTAPRDSASLP---PAFAST-----FYRPPRPAASSPCA-- 613
 QY 661 LHPEGLGLTFP--PFSQGHIMESANPPCEGLTYRTWSTSGFSSDSP-----P 708
 Db 614 -QPTTAAPRPTIEPTORAHQNSDLAESSTSEPPRPRIRSDMTSPAPVLFPEINSP 672
 QY 709 EAAPAPMAATPGL--PHSTPPVSDIWLVPPESEFOVD-----AAPVPAPD 753
 Db 673 RRPPLPAPLPDLEPAHTPPLS-----IPRODPTSDVPLMGSHLHSLPAPPTPLPS 728
 QY 754 PGLPBPVVLTPPPPPV-----HKPS--IPPSNRRLTYT 788
 Db 729 SOLPRLPRLNDPRLAIGPVLPEELHRRRYRPENTATFLTLRLSLPSNHLPOPTLN-CLLSA 787
 QY 789 YPDGARVYAGSLFESDCDMLVNA--SNPGHRRPG-----GLCH-----AFYQRFPE 832
 Db 788 VSDQTVSEHMLSELOTILPQSLNEENTNLGLSTEHLTALAHLYNQATVYSDRGP 847
 QY 833 APTPTFFIRREGLAATTLPRPIIAVADPYRYEONPKLEAAYRETCSRGTAANPLIG 882
 Db 848 LEPSPDTIKR-----IDTHTTGPSPHSPG-KRLGSGP---SAKGHSPDLIR 893
 QY 893-----SGIYQVSLSDFWEMRNHRPGDELY---LTPEAAMFWANKPAQVLTITTED 942
 Db 894 AKSEKVSQNY-LPFS-----FAHHNPISIAKHLISMKGEGVLSLDYSTGQRT 946
 QY 943 TARTANLALDATEVGRACAGCTISPG---IVHYOFTAGVGSQSKSIQO----- 992
 Db 947 GPTPKRIIQIDHYLDT-----NPKTTPVYHF---AGPAGGKTPYQQLLTKTK 994
 QY 993 -GDVNVVYVYTRRLNSMR---RGFAA--FTPTAARTYTRRYVIDEAPSLPHLLI 1045
 Db 995 FKDFR-VSCPTELRLREMTAMELHGSQSRFTWESSILKSSRIIVIDEIYIMPGYLD 1053
 QY 1046 LHMORASSYHL---LGP-----NQIPADIEHAGLPAIRPELAPISWMX 1088
 Db 1054 LSTLADPALLELYIILGDPLOGEYHSQSKSSNHRPS---ETLRLLPYI---DMTCWM- 1105
 QY 1089 VTHRCPADVYCELL-----RGAPYKIQT-----TSRVLSRLFWNEPAIGOKLVXT 1132
 Db 1106 -SVRIPOCIARLEQIHFSNANMGVIGSVTPHDQSPVLINSHASLSLTFN--SLGYR----- 1158
 QY 1133 QAKAANPCALIVHEAQATFETTTIATADAGLIQSSRAHAVALTR----- 1181
 Db 1159-----SCRTSSQGLTFCDPALIIVDNTKMLSS--ANGVALVATRSRSGVQPMGPS 1207
 QY 1182-----HTEKCVILDA--PGLAREGISDIYVNNFLAGVGCXNHR-- 1220
 Db 1208 STYVGCTNGSSAMFSAFNNSLIIMDYRFSLEPQL---KLITSPLTTRPKLNGATPSAS 1264
 QY 1221-----SVLRPGNDONLGLQAFPPSCQISAVHQAEBL 1254
 Db 1265 PTHRSBNPHLPRIHPLSYRDQEVYVNSTLPDQGPETRLDT--HFLPRLPLHFDLPPAI 1322
 QY 1255 GHRPAPVAALPDPCELEGLLYMDELTVSQVLFEL-----TDIVHGRMAAPS-- 1305
 Db 1323 --TPPVSVSYDP--PQAKASPYV--PGEE--PDSIAAFILPAIDPSTRRLHNDQSSNQRP 1376
 QY 1306 --QRAVAVSLTVGRYGRKRYLEAAH-----SDVRESIARFPTIGVQATT--- 1350
 Db 1377 WEDRFESLSC-----QPSLSLSAKHAPNHDPDLLPASINKRL-RFRPSDSPHQITADOV 1429

QY 1351 -----C-----ELYELVAMEKGGDSGAVLELDCNBDV----- 1381
 Db 1430 VLGLQLEFHSICRAVSRQPNSTVPFNPFLFACISLNEVQAQLSKSTQSTIYANASRSDPW 1489
 QY 1382 ---RIFFQKXCKKPTTGTETIANGKYGGGSGISAMSKFPCALFGR--WFAIIEKILLALP 1435
 Db 1490 RHTTVKIFAAQHKVNDGSIJFGSKACQTLALMDHYVILLGVKKRYQRIEFDN--ADRP 1546
 QY 1436 PNIF-----YGDAYEESVFAAASGASGCMVFENDESPEDSTQNNPSLFCVME 1486
 Db 1547 PNIIYHCGKTPNDRMCOEHL-----THSTPKIANDYTAPDQSGHGSVYLEALKMK 1599
 QY 1487 ECGMPQMLIRLY-HLVRSANILQAPKESILGFKPKHSGEGTLLMNTVMNMAIIACYEF 1545
 Db 1600 RLNIPLSHLIQHLVHKTNTSTQCPILCM-----RLTGEQGYDNDMDYNAVITYSQDYV 1654
 QY 1546 RDRFVAAFKGGDSVVLCSDY--KQSRMAAALLAGCGLKLVKY----- 1586
 Db 1655 GSCPIWV-SGDDSLI---DHPDPTRHDPWSVLKRLHREFLELTSHPLFCGYVYGAGCI 1710
 QY 1587 -RPIGLYAGVVAAPGLGTLPD-----VVRP--AGRLSEKMWGPGPRABQLRLAYCDF 1636
 Db 1711 RNPLALFCKLMIADVDDALDRLRLSTLTFTTGHLLGESLMLHPETHVQYSACFDF 1768

 RESULT 10
 RPO.ORSVC
 ID RPO.ORSVC STANDARD: PRT: 1612 AA.
 AC P89659; 039640;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [CONTAINS:
 DE METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
 OS Odontoglossum ringspot virus (isolate Korean Cy) (ORSV-Cy).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
 OX NCBI_Taxid=138661;
 RX
 RP SEQUENCE FROM N.A.
 RA MEDLINE=96381046; PubMed=8789059;
 RA Ikegami M., Isomura Y., Matsumoto Y., Chatani M., Inouye N.;
 RT "The complete nucleotide sequence of odontoglossum ringspot virus (Cy-
 RT 1 strain) genomic RNA.";
 RL Microbiol. Immunol. 39:995-1001(1995).
 CC -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
 CC RNA REPLICATION.
 CC -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
 CC CAPPING AND AN RNA HELICASE.
 CC -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
 CC BETWEEN CODONS FOR LEU-1112 AND GLN-1114.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL, D13941; BAA21828.1;
 DR EMBL, S83257; BAB949498.1;
 DR InterPro: IPR001788; RNA_dep_RNApol2.
 DR InterPro: IPR002588; Y_methyltransf.
 DR InterPro: IPR000606; Viral_helicase.
 DR Pfam: PF00978; RNA_dep_RNApol2; 1.
 DR Pfam: PF01443; Viral_helicase; 1.
 DR Pfam: PF01660; Y_methyltransf; 1.
 KW Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
 FT CHAIN 1 1612
 FT NP_BIND 826 833 METHYLTRANSFERASE/RNA POLYMERASE.
 FT ATP (POTENTIAL).
 FT SEQUENCE 1612 AA; 183252 MW; 3F803A2B9611ED0F CRC64;

Query Match 3.8%; Score 339.5; DB 1; Length 1612;
 Best Local Similarity 19.2%; Pred. No. 8.7e-14;
 Matches 331; Conservative 232; Mismatches 659; Indels 585; Gaps 85;

94 LHCF-----LRPGRDQRYNSAPTRGPANCRSALRGIP-----PADTYC 137
 123 VHCMPNLDIRVARRHINODTVSTY--LARLEKSK-RGLPVRQOQAFNKYMDDAVCC 179
 138 FDFGSRCAFAET-----GVALYSLHDLMPAD-VAEAMRHGXTRLAALHPREVLPP 191
 180 DKRPQDQSVLDLPKTYAVALHSIYDI-PADERGALLKRDVHICAAHFSNLL-- 236
 192 GTHTTSTYLL-----IHGDRAVVTYEGDTAGYNHDSILRAMITRT-----KIVG 238
 237 ---ETTSAPDELIGATPFYKSGDRLSEFFQNESTLNTYHSKYNVIKYCKFFPASNRFVY 293
 239 DHPFLYERVARIGHFVLLLT-----AAPSPMPYRP----- 271
 294 HKRPCTRTVMTFCCKTKVVTYFLFKGVYTRGEDSEOFYAMDEAMEYKKTMLMKCERT 353
 272 -----YRSTEVYRSIF-GRGSPSLFPSSACSTKSTF-HAAPHVIMDRMLFG 318
 354 IFDRRAAVNFMFKYKMDVITVPLEDGSVTSGMKMRSEVMNKKDEVYTVLNI----- 405
 319 ATLDQAFCCSRLMTYLRGISYKYVYGAALVANGMNASEDALTAATVITLCHQRYLR 378
 406 RTYQDKALTYKNVLSFEVESIRSRVILINGVSAREMDVKSVLQA-LSMTFL-----LQ 457
 379 TQMSKGRRLGVEHOKF---ITRLYSMLFEKSGRDYITGRQQLFAQCRRLWSLAFHL 435
 458 TK-LAEKKDOVVLAKFKPKEDVTNLEKWOISDAVGLPFSIK-----ETLISGPF-- 507
 436 DPRVLVDESVPCCRFTFLKVKAGFCFCFMRLGQECTCELEPAEGLVGDHGNMEAYEG 495
 508 ---VKVAEGLQI-----KTDPDEYITF-----ADKLIME-----YKA 536
 496 SEVDPAEPALDVSGITYAVHGHQLEALYRALNPQDIARAARLTATVELVASDRLSCR 555
 537 TE-----ELQHLDISKPL-----ERAEKYNNAL-----SELSTVLRKE-CDEFDIIOFK 577
 556 TYLGNKTFRTTY-----DGAHLEANGPEQVLSFDASRQSMGASHLLEYELPA 606
 578 NICEKHQIDDDVAKVIALMKNELITPFRKNPTREALS-DA-----LSLPLK 623
 607 GLOVRISNGLDCTATRP-----PGAPSAAPGEVAFCASALRYNRFTQHSILTGGLW 660
 624 DLMRFDLTLKLTSCAPRPPSVKTLDSGLLPKQSYGDERQF-----ESQSVASVD 672
 661 LHPEGLLITPPRPPGHIMESANPFCGEGITLYRTMTSGFSSDFSPREAAAPAMATPG 720
 673 FHLKSV-----BSVKMSMSAAVYTGRLKVVQOKNMY--DYLSASISAT-- 714
 721 LPHSTPVSIDWLPPESEEFQVDAVPRAPDAGLPGVYVLRPPRRPVHKPSIPPS 780
 715 VSNLCVKLKDVYGDAPESAE----- 734
 781 RNRRLTYTPDGAQVYAGSLFESDCDWLVNANSNGHRPGGGLCHAFYQRPPEAFYPTFEI 840
 735 --KSGVYDVVKG-----KWLTKPK-----GKCHAM----- 757
 841 MRGLAAYTLTPRPIIAVAVADVVEONPKRLAAYRETSRRTAAYPLIGSITYOVV 900
 758 ---GVALNNGEKVIV-----LLEMA-----DGFPIIGD----- 783
 901 SLSDAERNHRHPRDELYLTE--PAAMWFANKPAQOVLTITTEDTARTANLALIDATE 958
 784 -----WRRVAVSDSLIYSDMKLQTLUSCLKDGEVY----- 815
 959 VGRACAGCTISPGIVHYQTAGVPGSGKRSIOG--DVDVVVVPTRELNSWRRRG-- 1013
 816 -----PSDAKVTIVDVGVGCGTKTEILETVNFDEDLILVPGKEAKMIITKRANK 864

1014 -----FAFTPHTAARVTIGRRVYIDEAPSLPHLLH-----MORA 1051
 865 SGHVRATKDNVRYTDSFLMLHKPTY--NKLFIDEG-----LMLRTGCYNLIALSHC 915
 1052 SSVHLGDPNOIPAIID-----FEHAGVPAIRPLAFTSMWXVTHRCPAD 1096
 916 REAMVEDTQIPIFINNVANFPYKHFATLYYDIR-----EVR-----RLSLRCPAD 962
 1097 VCELIRGAYP-KIQTTSRVLRSL-----FVN--EPAIGKLVYQOAKA----- 1137
 963 VTHRMNSKYDGKVLCTNDVIRSVDAYRGKGVNPKSRPLKGIITFOSDKKELKERG 1022
 1138 -----ANGCAI-TYHEAGATTTETII-ATPADAGLIQSSAAHIVLFRTER-----CV 1187
 1023 YEEVSTGEINFTYHEIGQEFEDVSVRPLPTPLELISKSSPHVALVATRTKSEKYYSV 1082
 1188 ILDAPGLLEVG-----ISDVIVNNEFLAGEVGXHRPSVIPRGPNQNLGTQAFPPSCQ 1243
 1083 VLDP--LVKCSDSLKSXSDFLDMKYKADAGILXQLOVGSIFKG--EVL-----FVP-CP 1131
 1244 ISAYHOLAEIGHRPAPVAAVLPCCPEL--EGLIIV--POELVSDSVLVELTDIVH 1238
 1132 KSGY-----ISDQFYFDTLPGNSTILNEYDAVTMLRENNLNVKDCITDFSKS----- 1181
 1299 CRMAAPSORKAVLSTLVGRYGRRTKLYEAHSDVRESLARIPT--IGPVQATTCELYEL 1356
 1182 --VSVPQOQEEFPVIT-----RTAERPRSGLENIIVAMIKRNPSPDLTGILDEDT 1234
 1357 VEAMVERGODSAVLELDICNDVSRITFEOKC--NKFTTGETIAHGKVG----- 1405
 1235 AELVYKFWDAYITIDELSGGVNTPMTSDAIFRMMAKQEKSTIGLADPDFDLDAIDQYK 1234
 1406 -----OGISAKSTFCALFGPWRFAIEKEITALLPNI--IFYG 1441
 1295 HMIKAPQKQKLDLPODEYALQTIYVHSKQINAFGLFSELTROLLERIDSKFLFYT 1354
 1442 DAYEESV--FAAASGAGSCWVFENDSEPTQNNNSLGEVYMEGCPQWMLIRIYH 1439
 1355 RKTPEQIEEFSDSDSVYPMVEVLDELDSKYDSKQNEFCAVEALVWEVLGNGFL----- 1409
 1500 LVRSAMILQAPKESLKG-----WKHSGEPGLTLMNTVNMMAIIACHYEFRRFRV 1550
 1410 --EEVMQGHKRTSLKQVTAIGTKICVMQKRSQGVTFEIGNTVIAACLASIMPMKYIK 1467
 1551 AAFKGDSDVYLC--DYRGRNAALTAGGGLKLYKDYRP-IGLIYAG-VYVAGELGLTP 1605
 1468 AAFCGDDSMLYIPKGLDLPDIOSGANLMMNFEAKL--YRRRYGFCGRYIIHHDRAIV 1524
 1606 --DYVRFAGRLSEKNMGVGPERRAOLRLVACDFLRGLTNVQV--CYDVVASRYGVSPGL 1661
 1525 YVDVYKLISKGCKHI-KSLDHLIEFRISLDDVNASLNCAYIYQOLDATAIEVH----- 1577
 1662 VHNIGMLQTIADGKAHFTETIKPVL 1668
 1578 -----KTAVNGSFAFCSTIVKTYLSD 1596

RESULT 11
 RPO_TOML
 ID RPO_TOML STANDARD; ' PRT; 1616 AA.
 AC P03587; 041352;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [CONTAINS:
 METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
 OS Tomato mosaic virus (strain L) (TOMV) (TMV strain tomato).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
 OX NCBI_TaxId=12252;
 RN [1]
 RP SEQUENCE FROM N.A.
 MEDLINE=85157522; PubMed=6549393;

RA Ohno T., Aoyagi M., Yamanashi Y., Saito H., Ikawa S., Meshi T.,
RT Okada Y.,
RT "Nucleotide sequence of the tobacco mosaic virus (tomato strain)
RT genome and comparison with the common strain genome.";
RT J. Biochem. 96:1915-1923(1984).
CC -1- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC RNA REPLICATION.
CC -1- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC CAPPING AND AN RNA HELICASE.
CC -1- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC BETWEEN CODONS FOR GLN-1116 AND GLN-1118.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X02144; CAA26085.1; -
CC EMBL: X02144; CAA26082.1; -
CC PIR: A04195; WMTM8T.
DR Interpro: IPR001788; RNA_dep_RNApol2.
DR Interpro: IPR002588; V_methyltransf.
DR Interpro: IPR000606; Viral_helicase.
DR Pfam: PF00978; RNA_dep_RNApol2; 1.
DR Pfam: PF01443; Viral_helicase; 1.
DR Pfam: PF01660; Vmethyltransf; 1.
KW Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT CHAIN 1 1616 METHYLTRANSFERASE/RNA POLYMERASE.
FT NP_BIND 833 840 ATP (POTENTIAL).
FT SEQUENCE 1616 AA; 183564 MW; A8EC8929B5CF7CAFC C64;

Query Match 3.68; Score 322.5; DB 1; Length 1616;
Best Local Similarity 19.2%; Pred. No. 1.1e-12;
Matches 349; Conservative 205; Mismatches 660; Indels 601; Gaps 79;

QY 91 PNVLHCFGLRPVGR--DVQWYASAPTRG--PANCRSAL-RGIPRADRYTCDFGSRCA 145
DB 128 PNVLDVMDHNEGOKOSIELYLSRLERGNKHPFOKEADRYAEMNEVYVCHDTFOTCR 187
QY 146 FAEE--TG-----VALYSIHLMPAD--VAEAMARHGXTRLAALHLPEVLLPPG--TYHT 196
DB 188 HSECECTGRVYALAHSHIYI--PADEGALLKKNVYCAAFHFSENLLDESHVWDE 246
QY 197 TSYLLHHDGRAVYVEGDTSGAGYNDVSLIRAMIRTTKIVGDHPLVIERVRAIGCHEVL 256
DB 247 INACFORDGRLTFESFSESTLNSYSNLIKVCYCKT----- 284
QY 257 LITAAEPSPMPVPPRSTREYVVRSLFPGGSGSLFPSACSTKSTTHAVPIHIMRLML 316
DB 285 -----YFP-ASNEVYMK-----FLVTRVMTW----- 306
QY 317 FGAITLDDAFC--CSRIMTYL--RGISKYVGLALVANEAGNASEDALTAITATYALITICH 373
DB 307 -----FCKRSRIDTFLLYKGVAHK-----GVDSEQFKKAMEDAMHYKKT---LAMCN 350
QY 374 -ORYLTQALISG-----MRLGVEHAQKFTIRLYSWLFEKSGR-----DIYPRQ 418
DB 351 SERILLEDSSSVNYWEPKMDM-----VIVPLFDSLETSKRTRKREVLVSKDFV----- 399
QY 419 LQFYACR--RMLSAGPHLDPRVLFDESVPSCRTFLKKNVAKFCFCFMWLOEGECFL 476
DB 400 ---TYVLNHIRYQAKALYSNVLSFEESI--RSRYIINGVYAR----- 438
QY 477 EPAEGLVGDHNDNEAYEGSEVDPAPPAHLDVSGTYAVHSHOLEALYRALNVPQDIARA 536
DB 439 -----SEMDVDSILQSLSMTFPLH----- 458
QY 537 SRLTATVELVYASPDRLCETRVLGNKTPRTTVVDGAH----- 572

DB 459 -----TKLAVKDDLISKALGPKTVSOHWMEISLAFGNAPSIKERLINRKLITE 513
QY 573 --LEANGPEQYVLSFD--ASRQSMGAGSHSL-----TYELTPA--GLQVRISNGLD 618
DB 514 NALFIRVPDLVYFHDVLVEYKMSVDMPLVDIRKKMEETBEKMNALSELVLSKNSKFD 573
QY 619 CTATFPFGAPSAAPGEVAAPFCGALYRNFRTQHSITGLGLMHPGLGI1PPFSPGHI 678
DB 574 VDV-----FSQMCQSL----- 584
QY 679 WESANPCGEGTLYTRKWSISGFSSDSPPEAAPMAAATPGLPHSTPPYSD--IWLPP 736
DB 585 --EVDPTAAKVIYAVVNSNSGLTFPEQPTKANVALAL-----QDEKASDGLAVYTSR 637
QY 737 PSEFQVDAPVPAPDPAGLPQVYVLPPLPPPVVHKSP1PPPSRNRLLYTDPGAKY 796
DB 638 DVEPSTKSGMARGELQLAGSDV-----PES-----STRSEIE 674
QY 797 AGSLFESDCMLVNASNPGHRPGGLCHAFYQRPPEAFPTFEITMRGLAAYTLTPRPPII 856
DB 675 SLEQFH-----MATFASLIHQ---MCSIYTGGLKQGMKNFI--DSLVAISLSAAVSNL 724
QY 857 HAVAPDYRVQBNPKRLAEAYRETCSRGRGTAYPLLSGIIQVPVLSFDAM--ERNHRPG 914
DB 725 VKLIKDTAAIDLETRQKFGVLDVASKR-----VLVPSAKNHAMGVETHARK 772
QY 915 DELIYTF-----PAANFPEANKPAPVLTITEDTARTANLLEIDAATVEYRACAGCT 967
DB 773 YHVALLEHDEFGIITCDNMRRVAVSSVYSDMAKRTLRRLIK--DGEHVSSA----- 826
QY 968 ISPGIYVHOPTAGVPSGSKRSISIQG---DVDVVVYPTRELRSNRRR---GPAETPH 1020
DB 827 -----KVLVLDVGPGGCKTKEILSRVNFEDLILVGRQAAEAKIRRAVNSGIIVATKD 880
QY 1021 TAARV-----TIG-----RRVVIDEAPSLPRLILLH-----MORASSVHLGD 1059
DB 881 NVTVVSFLNNYKSGARCOQKRLFIDEG-----LMLHTGCVNFIEMSLCDIAYVGD 933
QY 1060 PNCIIPALDEHAGLVPA--TRPELAPTSMMXVYTRCPADVCELRGAYP--KIQTSHVLR 1116
DB 934 TQOIPYINRYTGPPRYAHFKLEDEYETRTTLRCPADYTHFLNQRTEBHVMCTSEKK 993
QY 1117 SLFW-----NEPAIGOKLVYTOAKAA-----NPGATVHEAGATFTETII 1159
DB 994 SVQGWYSGAASINPYSKRLPKGILFTFQSDKEALLSRGADVHTVEVGEFYADVSLV 1053
QY 1160 -ATAADARGLIQSRAHAIVALTRETK-----CYILDA-----P 1192
DB 1054 RLTPPTVYSIILARDSPHVLVLSLSHHTKSLKYTVVMPPLVSIINDLERVSSYLDMYKVA 1113
QY 1193 GLIREVGISDIYVN--NPEFLATGGEVGMHRPVIIRGNDDMLGTLQAPSPSCQISAYHQLA 1251
DB 1114 GTOXQLOVDVSVEKFNFLVAPKVTG-----DISDQFYDKC----- 1150
QY 1252 EEIGHRPAPVAAVLPCCPELEOGLIYMPDELTVSDS--VYVFEITDI-----VHC 1299
DB 1151 -----LPGNSTILNNNDAYTKMLTJDSLWVKCILLMSK 1184
QY 1300 RMAAPSORKAVLSTLVGRYRRTKLYEAHSDVRESIARTPT--IGPVQATTCELYELV 1357
DB 1185 SVAAPKDVKPTLLPMV-----RTAAEMPRQOTGLENIYVAMIKRNFNSPELSGVYDIENTA 1239
QY 1358 EAMVVEKGODSAVLELDLCNRDVSRTFFQKCNKF--TTGETIANGVVG----- 1405
DB 1240 SLVYDKFFDSYLLKEKKRPKNFS--LFSRESLNKRAKOEQVYTIQGLADPFDVLDPAVD 1297
QY 1406 -----OGISAMSKTFCALFCFPMWRAIEKELLALL----- 1434
DB 1298 QYRHMIAKQPKQKLDLSIQREYALQTIYVHSKINAIFFPLRPSLRLDLSIDSSRFL 1357
QY 1435 -----PPNLT--FYGDAYEESVFAAAYSGAGSCMVFNENDSPSDSVONNNSLGECEYME 1486

CC	-1-	MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC	BETWEEN CODONS FOR GLN-1116 AND GLN-1118.	
CC		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC		
DR	EMBL; AJ243571; CAB62911.1; .	
DR	EMBL; AJ243571; CAB62912.1; .	
DR	InterPro: IPR001788; RNA_dep_RNApol2.	
DR	InterPro: IPR002588; V_methyltransf.	
DR	InterPro: IPR000606; Viral_helicase1.	
DR	Pfam; PF00978; RNA_dep_RNApol2; 2.	
DR	Pfam; PF01443; Viral_helicase1; 1.	
DR	Pfam; PF01660; Vmethyltransf; 1.	
KM	Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.	
FT	CHAIN 1 1616	
FT	CHAIN 1 1116	
FT	NP_BIND 833 840	
FT	ATP (POTENTIAL)	
SQ	SEQUENCE 1616 AA; 183606 MW; 4BF9913CE79ADE33 CRC64;	
Query Match	3.6%; Score 320.5; DB 1; Length 1616;	
Best Local Similarity	19.3%; Pred. No. 1,5e-12;	
Matches 350; Conservative 223; Mismatches 646; Indels 591; Gaps 83;		
QY	91 PNVLHRCFLRPVGR-DYQRWYSAPTRG--PAANCRRSAL-RGLRPADRTYCDFGSFCA 145	
Db	128 PNLDPVRIMRREGKDSIELYLSRLERGNKHVPNFOKAEFRVAEMPEVEVCCHDTFCOR 187	
QY	146 FAAR--TG---VALSYHLDMPAD-VAEAMARRGHXTRLVALMLRPVELLPFG--TYHT 196	
Db	188 HSQCCTGYRYAIALHSTYD-PADFEQAALLRNHVCHTAAFHSSENLLEDSHVINDE 246	
QY	197 TSYLLIHGDGRAVVYEGEDTSAGYNHVSILRAMITTKIVIGDPLVIERYVRAIGCHEVL 256	
Db	247 INACFORGDRLTFSPASESTLYNSHYSNILKYVCKT----- 284	
QY	257 LLTPAAPERPMPYPYPRSTEVYVRSIFPGSGSLERSACSSTSTHAADVNHVHWDLML 316	
Db	285 -----YPP-ASNREVMYKE-----PLTVRVNW---- 306	
QY	317 FGALIDDAQFC-CSRLMTYL--RGISTKYGVGALVANEG-WNASEDALTAIXTYAAULTIC 372	
Db	307 -----ICKESRIDTEPLLKYGAANK-----VDNQEFKAEDMAHMYKKT---LAMC 349	
QY	373 H-QRYLTQAIKSG-----MRRLGEHAKOFTIRLYSMLEEKSGR-----DYIPGR 417	
Db	350 NSERILLIEDSSSVNYWPKMKRDM-----VIYPLFIQSISETSKRTREVLYVSKDFV-- 399	
QY	418 QLOFYAGCR--RWLSAGFHLDPRVLYVDESYPGCCRFPLFKVAKPCCFPMMLIQDECTCF 475	
Db	400 ----YYLVNHRITYQAALTYINVLSTVESI--RSRYIIINGVTAR----- 438	
QY	476 LEPAEGLVGDHNDNEAYEGSEVDPAEPAHILDVSGTYAVHGHOALEALYRALNWPDIAR 535	
Db	439 -----SEMVDVKSLLQSIISMTPFLH----- 458	
QY	536 ASRLTAVELVASDRLECRTVLGNKTRPTTYVDGAILLEANGPBQIYVLPDASQSGAG 595	
Db	459 -----TKLAVLKDDLLISKFPALGPKVSOHWMEISLAGN-----APPSIKERL-- 503	
QY	596 SHSLTYELTPAGLOLRSSNGLDCTATFFPGGAPSAPAGEVAARCASYLRNRFTQHSL 655	
Db	504 INKLRIKITEALEIRP---DLVYTF-----HDLRVSEYKM 537	
QY	656 TGGIWL-----HPGGLIGIPPPSPGHIMESANPFCEGTLTYRTWTWGTSGFSDSFSPA 710	
Db	538 SYVMPVYDIKKMETEEENNALSELVLTSDPEFDV-----VFQOMCOSILEVDPTA 591	

QY	711	AAPAMATPG-----LPHSPVSDIIVLPSPSEEQVDAVPAVPAPBAGLPGVYLP	765
Db	592	AKVIAVAMSSESGLTTFEEQPTFANALALDSEK-----ASD-----GALVLT-	635
QY	766	PPPPVPPSPISPPPSNNRRLLYTPPDGAVVAGSLTESDCDWLVNASNPGHPPGGGLCHA	825
Db	636	-SRDVEEPI-----RGSNARGILO-LAGLS-----GDYES	665
QY	826	FYQRFPPAPYPTFEFIMREGLAATLTPRPILH-----AVADYVEONPKRLEAAVET	879
Db	666	SYTSEIEESLEQFHM-----ATASSILHKOMCSIYTPGLPYQOMKNEIDSLVSL	717
QY	880	CSR-----RGTAVPL-----LGSYIYVPSLSFDMW--ERNHRPDELYL	919
Db	718	SAAVSNLVKLIKLTQALIDETROKFGVDVASKRMLVYKSAKNHAGVETHARKYHAL	777
QY	920	TE-----PAAMFEPANKPAQVPLFTEEDTARTALALEIDAATVEVGRAGCGTISPI	972
Db	778	LEHDEBITTCDDMMRRAVAVSSSESVISDAKLTFLRLIK-DGEFHVSA-----	826
QY	973	VHVQFAGVBSGKSRSIQOG---DVDVAVVPTRELNSMRR---GFAEPTPTAAV	1029
Db	827	KVLYDGVGVGCKTEILSRVFEEDLLVYGRQAEHRRRANAISGIVATKNVATY	885
QY	1026	---TG-----RRVIDEAPSLPHELLH-----KORASSVHLIDPNQIP	1064
Db	886	DSFLMNGKARGCQFRFLFDEG-----LMLHGTGVNFIEMSLCIDAIYVYGDYQIP	938
QY	1065	ALDEHAGLVPA--IRPELAPTSMWAVYTHRCPADVCELLRGAYP-KIOTTSVLSLFW-	11220
Db	939	YINRVTEGPPYPAHPAKLEDEVEYTRRTTLRCPADVTHFINORYEGHVMCTSSKKSVQSE	998
QY	1121	-----NEPATGOKLVXVQAAKA---NEGATVIEAOGAFPTETIT-ANAD	1163
Db	999	MVSAASINVSRLPKGKITLTFQDSKEALLSRGADVHTVHEVOGETAYDASVLRPLPT	1056
QY	1164	ARGLISSRAHAIVALTRHTEK---CVILDA-----PGLIRE	1197
Db	1059	PVSLIARDSPHVLYLSRHTKSLKYTYVAMPDVLSTIIRDLERVSYLDMYKVDAGTOXQ	1118
QY	1198	VGISDIYVN-NFPLAGEVGNHNPSTYIPGNPDQNLGTIQAPPPSCQISATYHQLAEELGH	1256
Db	1119	LOYDSVEKRNLEVAAPKTG-----DISDQFYDKC-----	1150
QY	1257	RPAPVAVLPCEPELEGLLYMPQELTVSDS--VLVFEETLDI-----VHCROMAP	1304
Db	1151	-----LGNSTLNNDAVYMKLTDISLVNKCIDLMSKSVAP	1188
QY	1305	SORKAVYLSTLVGRYGRKTKLYEAHSDVESLARPIPT--IGPVQATTCLELVEAWE	1362
Db	1190	KDVAPFLIPWV-----RTAEMRROGLLENVLAMIKRPNFNPBELSGVVDIENTASLYVD	1244
QY	1363	KGDGSAVLELDCNRDVSITFEQXXCNKF-TTGETIAGKVG-----	1405
Db	1245	KFEFYSYLKEKRRPNKNFS--LFSRESLNRMIAKOQOYVITQGLADDFVDVLPAYOYRHM	1302
QY	1406	-----QGISAMSKFPCALFGWFPAITEKILL--	1433
Db	1303	IKAAPKOKLDLSIOTEXPALQTVYHASKIMAFGLFSELRLOELSDSRFLFTFRK	1362
QY	1435	-PPPI--FYGDAAVEESVFAAAGSAGSCWVFENDFEFPSTONNSLIGECVWMECGMP	1491
Db	1363	TPAOIDEFPED-----LDSHPMDVLELDYSKYKSONERHCAVEYIWRGLIE	1412
QY	1492	QMLIRLVHVSAMILQAPRESLKGf-----WKHSGEPTGLTNTVNNMAIIAHC	1542
Db	1413	DLFALEV-----WKQGHRTKTLKDTAGIKTCIMVQKRSQDVTEPTIGTIVIIASCLASM	1466
QY	1543	YEFDFPVAAVAFKDDSVYL-----CSDYROSRAAALIIAGCGLKL-KVDYRPIGLYAGVY	1597
Db	1466	LPMEKLLKGFAGCDSDSLFLYPPKGC--GYRPIQOAAINMNFVAKLTKKQY--GYCGGRV	1521

```
QY      1598 A - PGGTLPDVYVRFAGRLSE---KNWGGPPEARQLAVCDPLRGITNNAQCVC--D    1649  
                :   |       :   |   |   :   |   |   :   |  
Db      1522 IHHDRGCIIVYYDDPKLISKLGAKHIND-----DHLEFRSLCDVAESLNCAVTTOLD    1577
```

```
QY      1650 VVSRYGVSP     1659  
            :   |       :   |  
Db      1578 AVGEVHTAP     1587
```

RESULT	14				
RPPO_CRAW	RPPO_CRAW	STANDARD;	PRT;	1597	AA.
AC	066220;				
DT	20-AUG-2001	(Rel. 40, Created)			
DT	20-AUG-2001	(Rel. 40, Last sequence update)			
DT	20-AUG-2001	(Rel. 40, Last annotation update)			
DE	RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (182 KDA PROTEIN) [CONTAINS:				
DE	METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (125 KDA PROTEIN)]				
OS	Chinese rape mosaic virus (CRAW) (oliseed rape mosaic virus).				
OC	Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.				
OX	NCBI_TaxID=42007;				
RP	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=96197410; PubMed=8616237;				
RA	Aguilera I., Sanchez F., Martin-Martin A., Martinez-Herrera D.,				
RA	Ponz F.;				
RT	"Nucleotide sequence of Chinese rape mosaic virus (oliseed rape mosaic				
RL	virus)", a crucifer tobamovirus infections on Arabidopsis thaliana.";				
RL	Plant Mol. Biol. 30:1191-197(1996).				
CC	-1- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL				
CC	RNA REPLICATION.				
CC	-1- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA				
CC	CAPPING AND AN RNA HELICASE				
CC	-1- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS				
CC	BETWEEN CODONS FOR GLN-1103 AND GLN-1105.				
CC	-----				
CC	THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION				
CC	CC Between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	CC the European Bioinformatics Institute. There are no restrictions on its				
CC	CC use by non-profit institutions as long as its content is in no way				
CC	CC modified and this statement is not removed. Usage by and for commercial				
CC	CC entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	CC or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; U30944; AAB60599.1; -				
DR	InterPro: IPR001788; RNA_dep_RNAPol2.				
DR	InterPro: IPR002588; V_methyltransf.				
DR	InterPro: IPR000606; Viral_helicase.				
DR	Pfam: PF00978; RNA_dep_RNAPol2; 1.				
DR	Pfam: PF01443; Viral_helicase1; 1.				
DR	Pfam: PF01660; Vmethyltransf; 1.				
KW	transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.				
FT	CHAIN 1 1597				
FT	CHAIN 1 1103				
FT	NP_BIND 823 830				
FT	VARIANT 919 919				
FT	VARIANT 1286 1286				
FT	VARIANT 1597 AA; 181621 MW; 7256A908BD3308F2 CRC64;				
FT	SEQUENCE				

```

Db 934 PYPAHEAKLVVDEKEDR-----RVTLRCPADVYFFLNOKYDGSVLTSSVSERSVSAEV 987
OY 1123 -----PAIGOKLVYVTOAK-----AANGALITVHAGCAFTETITII-ATADAR 1165
Db 988 RKGALNPITLPLEGKILITTOADKFEELDKGYDVNTVEHVOGETEKTALVATLTP 1047
OY 1166 GLIOSSRAHAIVALTTRTEKC-----VILD-APGLIREVG-ISDVIY-----1205
Db 1048 EIIISRAAPHVVALTRHTTCKKYTVVLDPMVNVISLGLTNSLLEMYVESGTQXQ 1107
OY 1206 -----NNFFLAGGEVGHXR-----PSYIPKGNP-----DONLGLQAFPPSCQ 1243
Db 1108 IDTFVGTNLFFVPPKSGDWRDMQFYDILLP-GNSTILNEFPAVTNNLMDISLNVKDCR 1166
OY 1244 I--SAYHQAELGLHRA-----PVAALVPCPELEGGLLYM-POELTVSDSVLVELT 1294
Db 1167 IDSRSKVQVPE---RVEFKPKLRTAEMPRTAGLLENVAMIKRMNAPDLTGTDIE 1223
OY 1295 DIYHCRMAAPSQRKAVLSTLVGRYGRRTKLYEAAHSD---VRESLARFT---PTIGP 1345
Db 1224 D-----TASLVVEKFWDAVYVKEFSGTDGMAMTRRESFSLKQESSVYVQ 1269
OY 1346 VQ-----ATGCEIVELVAMVEKGGDSAVILEDCNDRDVGRIFFQKCKKFTTGER 1398
Db 1270 LAFPNVYDLPAYDEYKHMKSQPRQ-----KLDLSIOD-----EYPALQT 1309
OY 1399 IAHKVGOGISAMSKFCALFCPFWFAIEKEIILALPPN--IFYGDAYESV--FAAAS 1454
Db 1310 IYV-----HSKRIALFQMFSELTMLLETIDTSKFLFYTRKTPQIDEEFEDLD 1360
OY 1455 GAQSCWAFENDESEFDSTONNESLGLCEVYMEGCPQWLIRLYHLVRSAMIIQAPRESL 1514
Db 1361 SSQAMELLEDISKYKQSQNEHFCAYEYKIMWKIGIDMDLAEV-----WROGHKRTTL 1413
OY 1515 KGF-----WKHSSEGTLLMNTVMMAIIAHCEYERDPRVAFKGDSDSVLCSY 1565
Db 1414 KDYTAGIKTCWYQKSGSDVTTTIGNTIIIAOLSSKIMPDKYIKAFCDSDSLIYIP-- 1471
OY 1566 ROSRNAALTAGCGL-----KLKVDYRPIGLYAG-VVYAPGLTLP--DVYRFAGRISEKN 1618
Db 1472 -KDLDPDIOGANTLWNPFAKLFRRKKYGFYCGRYVHHHGRGAIYVYDPLKLSKLGCKH 1530
OY 1619 WCGPERABOBLAVCDFELRGLTNVAOCV--DVYSRVYGVSPGLVHNLGLQTLAD 1674
Db 1531 I-RDEVHLEBELRSLCDVSNLNNCAVFSQDEAVAEVHEHTAVGAFVYCSIIKYLSD 1587

RESULT 15
RPO_TM0B
ID RPO_TM0B STANDARD; PRT; 1616 AA.
AC P90211; O83484;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [CONTAINS:
DE METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
OS Tobamovirus OB.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
OX NCBI_TaxID=31749;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=93389450; PubMed=8376970;
RA Ikeda R., Watanabe E., Watanabe Y., Okada Y.;
RT "Nucleotide sequence of tobamovirus OB which can spread systemically
in N gene tobacco.";
RL J. Gen. Virol. 74:1939-1944(1993).
CC -I- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC RNA REPLICATION.
CC -I- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC CAPING AND AN RNA HELICASE.
CC -I- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC BETWEEN CODONS FOR ALA-1115 AND GLN-1117.

```

```

CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.slb.ch).
CC -----
DR EMBL: D13438; BAA02700.1; -
DR EMBL: D13438; BAA02701.1; -
DR InterPro: IPR001788; RNA_dep_RNApol12.
DR InterPro: IPR002588; V_methyltransf.
DR InterPro: IPR006066; Viral_helicase1.
DR Pfam: PF00978; RNA_dep_RNApol12; 1.
DR Pfam: PF01443; Viral_helicase1; 1.
DR Pfam: PF01660; Ymethyltransf; 1.
KW Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT CHAIN 1 1616
FT NP_BIND 832 839
FT MP_BIND 1 1115
FT SEQUENCE 1616 AA; 182997 MW; 146BD679AA46575 CRC64;

Query Match 3.5%; Score 319.5; DB 1; Length 1616;
Best Local Similarity 18.1%; Pred. No. 1.7e-12;
Matches 329; Conservative 218; Mismatches 580; Indels 693; Gaps 74;

OY 133 ADRTYCGDGRSFC---AFAETGVALYSLHDLV--PAD-VAEAMARHGRTALYALHLP 185
Db 174 SDEVYCNNTFOCCESNRYSSGGRYVALSLSLVDIPADEGALRLRNHTVLYAAHFHAE 233
OY 186 EVLLPQGYH--TTSYLLIHDGRAVYVEGDSAGYNHDVSLIRAMIRTT-----KIV 237
Db 234 ELLEVTVELPRTIGGIFGSDGDKINCFSENESTLWNSHYSNLKTVCTTPPASRRY 293
OY 238 GDHPLVIERVRAIGCHPVLILTAPEBSPMPYVPYRSTEVYRSIFPGGSPSLFPSPAC 297
Db 294 YKKEFLITRVNWFCKFKTLDI-----YLY-----RGVYHRGCDQOEYSAM 336
OY 298 ST-----KSPFHAV-----PYHINDRLM-----LFGATIDDO----- 324
Db 337 EDAMHYKKTLLAMLSERIVLEHSSVNVWPPKKDMVYVPLFVLSLETQRTKREYIVSK 396
OY 325 -----AFCCSRMTFLRGISYVVTGALVANGMA----- 355
Db 397 DFVYTVLNHRTYQAKALYNNVLSFEVSTRSVIINGYARSEMDYDALLOSMAMTF 456
OY 356 -----SEDALTAYITAAVYTI-----CHORY-----LRTQATSKGMRL 389
Db 457 LITKLSMLKDELIVSKFTLSAKSVHEHVMDEIKRGGNMPPSLKESLIRKRLISGAEEL 516
OY 390 GVHNAKFTIRLYSMLEKSGRYITGRQL-QYACCRKWLSGPHLDPRVLYFDESV- 447
Db 517 EIEVPMYVTFHDFVAFKASVEMPTIDISKLSAESYSALSLS-----VLENSKDF 572
OY 448 -----CRCRFLKLVAGKFCFCFMRWLGQECTC---FLEPEGLVGHGHNEAYE- 494
Db 573 DLEKFSRMAINCVDNIAKIVAV--LSNEGTVLLPKEPEEGMAARMAKSGEDEV 630
OY 495 --GSEVDPAPPAHLVDVSGTYAVVHGQLEALYRALNVPODIAARASHLTATVLEVASPRL 552
Db 631 TLGSQDNT-----DUTSKSMVTSISLPLCGIASEL 661
OY 553 ECRTVLGKNTFTTVDGAHLEANGPEQYVLSFDSRSGAGSHSLYTELPAGIQVRI 612
Db 662 SCDTFVRNE-----EINSELEYHML--AAESYISNMKASIVYS--GPLYOQ 704
OY 613 SSGNLDCTATFPFGGAPSAPEV-----AAPCSALRYNRRTQRSISLNGGLLHPE 664
Db 705 MONYDLSLA-----ASLSTVSNLKLKLVDOSSVGFDLSLKGVFVRRK---MWL--- 752
OY 665 GLIGIPPPSPGHIMESANPFCGEGTLVTRTWSGFSDFSDFPEAAMATPGLPHS 724

```

```

Db 753 -----IKPTLKNHSGVQKFDGKCFLLALSY-----HN 781
QY 725 TRPVSDI-WLPPSEEFQVDAAPVPADPAGLPGBVLTTPPPPPVHKPSITPPSRNR 783
Db 782 ELPICDADMSKVAVSNESWV-----801
QY 784 RLITYTPDGAKVYAGSLFESDCDWLVNASNPGHRPGGLCHAFYQRRPEAFYPTETIMRE 843
Db 802 -----YSDMAKL-----808
QY 844 GLAAVLTTPRLIHAVAPDYRQNPKELEAAYRETCSRGRGTAAYPLGSGIYOVPSLS 903
Db 809 -----RYLKRSTIGEMPISTVS 823
QY 904 PDAMERNHRPDELVTLEPAANWFANPKPAOVLTTEDTARTANLALETDAATEVGRAC 963
Db 824 -----SAKVTL-VD-----831
QY 964 AGCTISPGIVHYQTAGVPSGSKRSIQG---DVDVYVYPTRELNRSMRRRGFAATPH 1020
Db 832 -----GVPGCGKTKETLRRVNESEDLVLPGRKEAAMIRKR--ANQSGN 873
QY 1021 TAAR-----VTIG-----RRVVIDEAPSLPHLL--ILHMORASSVHLLGDP 1060
Db 874 IVANNNDNKTVDSEFLMNLGKPGVCQFRKRLFYVDEGLMLHPCYVFLVLCNEAFVFGDT 933
QY 1061 NOIPAIIDE-----EHAGLVPAIRPELAPTSMWXYTHRCPADVCELIRGAYP-KIQTTSS 1112
Db 934 QQIPYINRVQNFPEPQHSKLIYDETEKRRT-----TLRCPVDYTHFLNOCYDGAVTTS 988
QY 1113 RVLRSLFW-----NEPALGOKLYXTQAK-----AANPGALTYHEAGATFTE 1155
Db 989 KTRRSVGLLEVYGAAYWNPVTKPKLKIYFTQSDKLTMLSRGYQDVNTVHEIOGETYEE 1048
QY 1156 TTII-ATADARGLIQSSRAHAIVALTFRTEKC-----VILD-APGLREYISDVIYNF 1208
Db 1049 VSLVRLTPTPIHILISRESPIVGLTRHT-RCKRYTYVLDPLVKLYRDL-----ECVSNF 1103
QY 1209 FL-----AGGEVGHXHRPSVIPRGNDNLGTLQAFPPSQOISAYHQLAEE 1253
Db 1104 LLDVYWDVSAXQLOVSGVYLAENLEFVQAPKSGDAQD--LQFYDKCLPGNSTVLNE- 1159
QY 1254 LGHRPAPVAALPPCELEGLIMPOELTVSSVLYFE-----LTDIVHCR 1300
Db 1160 -----FDVAVTMNCSDI-----SLNVKDCVLDFSKSVPLPRDNTKVPMTPI--R 1201
QY 1301 MAAPSOR-KAVLSTIVG-----RYGRRTKLYEAH 1329
Db 1202 TAERPRPSOGLLENLVAMIKRNFNSPELSTVDMENTASVADRFPDSTYFLKDKLSGCSL 1261
QY 1330 SDV-----RESLARI-----PTIGPVQATTCELYELVE-----AMVEKGODGSAY 1370
Db 1262 GDSGKNIIIDROALLIRMEKQEKSTIQQLAD-----YDFVLDLPAIDQYRHIKSPKO-- 1314
QY 1371 LEIDLGNRDVSRITTFQOKCNKFTTGFTLAHGKVGQISAMSKTFCALFGPWRPAIEKEI 1430
Db 1315 -KIDLSIQ-----SEYPSLOTIY-----HSKRINALFGPIFSELTROM 1352
QY 1431 LALLPKN--TFYGDVAYEESV--FAAAVSGAGSCWVFENDESEPTQONNFSLGEQVME 1486
Db 1353 LSAIDTSRYLFTFRKTPPEQLEEFFSDIDAHQPMHEVLELDVSKIDKSQNEFHCAVEYEIWK 1412
QY 1487 ECGMPQWLRLYLHLVRSAMTLQAPKESLKGf-----WKHSGEPGTLMTNTVWMA 1537
Db 1413 RLGIDEFLEAV-----WKQGRHRTKTKDYTAGIKTKLWYQRKSGDVTTFTGNTVITAA 1465
QY 1538 IIAHCYEFRRFYAAAFKGDSDSVYLCSDYRQSRNAALIAAGCL-----KLKV 1584
Db 1466 CMASMLPMEKVIYKAAFCGDDSLVY-----LPKCELPNIOSCANLMMNFEAKL 1513
QY 1585 DYRPIGLYAGVVA--PGIGTLPDVYVFAGRBSEKNWGPERRAEQRLAVCDFLRLGLT 1641

```

```

Db 1514 FKTYGYFCGRYVYIHDRCALIVYVDPKLIISKLGAKHI-TDKEHLEEFRLSLADVSKSLN 1572
QY 1642 NVAQVCY--DVYSRYGVSP 1659
Db 1573 NCAYYAQDDEAVREVHKTAP 1592

```

Search completed: March 7, 2002, 14:08:31
Job time: 279 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 7, 2002, 14:03:02 ; Search time 49.94 Seconds
(without alignments)
4973.376 Million cell updates/sec

Title: US-09-468-147-91

Perfect score: 9013
Sequence: 1 PGITTAIEQALAAANSALA.....FTETIKPYLDITNSIIOKRE 1698

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505.seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :

SPTREMBL.17:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.rodent:*
12: sp.virus:*
13: sp.vertibrate:*
14: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9001	99.9	1698	12	O9YLK3 hepatitis e
2	8812.5	97.8	1708	12	O9YK10 swine hepat
3	8780.5	97.4	1708	12	O9YLK1 hepatitis e
4	7730	85.8	1707	12	O9YLK2 hepatitis e
5	7475	82.9	1685	12	O9E8G6 hepatitis e
6	7468	82.9	1693	12	O9E8G6 hepatitis e
7	7457	82.7	1693	12	O9W1L5 hepatitis e
8	7456	82.7	1693	12	O9W1L5 hepatitis e
9	7424	82.4	1693	12	O8E4L4 hepatitis e
10	7412	82.2	1693	12	O8E4L4 hepatitis e
11	7333	81.4	1693	12	O8E4L4 hepatitis e
12	7329	81.3	1693	12	O9WC28 hepatitis e
13	7304	81.0	1693	12	O9WC28 hepatitis e
14	7301	81.0	1693	12	O8E4L4 hepatitis e
15	3444	38.2	727	12	O8E4L4 hepatitis e
16	2520	28.0	479	12	O7I1A4 hepatitis e
17	2365	26.2	452	12	O7I1A4 hepatitis e
18	2284	25.3	487	12	O8E4L4 hepatitis e
19	1821	20.2	491	12	O9YK12 hepatitis e

20	1368	15.2	290	12	O8E4L4 hepatitis e
21	1183	13.1	245	12	O8E4L4 hepatitis e
22	1081	12.0	225	12	O9W1K5 hepatitis e
23	1074	11.9	217	12	O8E4L4 hepatitis e
24	1025	11.4	212	12	O9YK11 hepatitis e
25	988	11.0	210	12	O9E4L4 hepatitis e
26	988	11.0	210	12	O9E4L4 hepatitis e
27	988	11.0	210	12	O9E4L4 hepatitis e
28	988	11.0	210	12	O9E4L4 hepatitis e
29	981	10.9	210	12	O9E4L4 hepatitis e
30	981	10.9	210	12	O9E4L4 hepatitis e
31	913	10.1	193	12	O9W1K6 hepatitis e
32	884	9.8	182	12	O8E4L4 hepatitis e
33	769	8.5	158	12	O9W9G2 hepatitis e
34	763	8.5	152	12	O8E4L4 hepatitis e
35	709	7.9	134	12	O9YK14 swine hepat
36	692	7.7	158	12	O8E4L4 hepatitis e
37	666	7.4	134	12	O9E4L4 hepatitis e
38	666	7.4	136	12	O9W8P3 hepatitis e
39	665	7.4	136	12	O9W8P3 hepatitis e
40	662	7.3	133	12	O9E4L4 hepatitis e
41	661	7.3	133	12	O9E4L4 hepatitis e
42	658	7.3	136	12	O9W9Y6 hepatitis e
43	654	7.3	136	12	O9W8P6 hepatitis e
44	646	7.2	123	12	O9E3S9 hepatitis e
45	646	7.2	123	12	O9DQV7 hepatitis e

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT; 1698 AA.
O9YLK3	O9YLK3		
AC	O9YLK3		
DT	01-MAY-1999 (TEMBREL. 10, Created)		
DT	01-MAY-1999 (TEMBREL. 10, Last sequence update)		
DT	01-JUN-2001 (TEMBREL. 17, Last annotation update)		
DE	POLYPROTEIN (FRAGMENT).		
GN	ORF1.		
OS	Hepatitis E virus.		
OC	Viruses: ssRNA positive-strand viruses, no DNA stage.		
OX	NCBI_TaxID=12461;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-HEV-US1.		
RC	MDLINE-98178637; PubMed=9519822;		
RA	Schlauder G.G., Dawson G.J., Erker J.C., Kwo P.Y., Knigge M.F.,		
RA	Smalley D.L., Rosenblatt J.E., Desai S.M., Mushahwar I.K.,		
RT	"The sequence and phylogenetic analysis of a novel hepatitis E virus		
RT	isolated from a patient with acute hepatitis reported in the United		
RT	States.";		
RT	J. Gen. Virol. 79:0-0(0).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-HEV-US1.		
RC	MDLINE-99190429; PubMed=10092008;		
RA	Erker J.C., Desai S.M., Schlauder G.G., Dawson G.J., Mushahwar I.K.,		
RA	"A hepatitis E virus variant from the United States: molecular		
RT	characterization and transmission in cynomolgus macaques.";		
RT	J. Gen. Virol. 80:681-690(1999).		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-HEV-US1.		
RC	Schlauder G.G., Erker J.C., Dawson G.J., Desai S.M., Knigge M.F.,		
RA	Kwo P.Y., Smalley D.L., Rosenblatt J.E., Mushahwar I.K.,		
RA	Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.		
RN	[4]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-HEV-US1.		
RC	Erker J.C., Schlauder G.G., Dawson G.J., Desai S.M., Mushahwar I.K.,		
RA	Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.		
DR	EMBL, AF060668; ADD15812.1; -.		

DR InterPro: IPR000606; Viral_helicase1.
DR InterPro: IPR002588; V_methyltransf.
DR InterPro: IPR002589; DUF27.
DR Pfam: PF01443; Viral_helicase1; 1.
DR Pfam: PF01660; Methyltransf; 1.
DR SMART: PF01661; DUF27; 1.
DR SMART: SM00506; A1pp; 1.
DR Polyprotein.
FT NON_TER
SQ SEQUENCE 1698 AA; 186069 MW; 1C670601150F4EA CRC64;

Query Match 99.9%; Score 9001; DB 12; Length 1698;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1698; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PCITTAIEQAALAAANSALANAVVPRFLSRVOTELINLMOPROLVPRPEVLMMNPOR 60
DB 1 PCITTAIEQAALAAANSALANAVVPRFLSRVOTELINLMOPROLVPRPEVLMMNPOR 60
QY 61 VHNLEEOYCARAGRCLEVGAPHSINDPNVLRHCELRPGVDQRYSAPTGPAAN 120
DB 61 VHNLEEOYCARAGRCLEVGAPHSINDPNVLRHCELRPGVDQRYSAPTGPAAN 120
QY 121 CRRSALRGLPADRTYCEDEGFSRCAPAEETGVALYSLHDLPADVAEAMARHGXTRLYAA 180
DB 121 CRRSALRGLPADRTYCEDEGFSRCAPAEETGVALYSLHDLPADVAEAMARHGXTRLYAA 180
QY 181 LHLPEVLLPCTGHTTSTYLLHGDGRANVYEGDTSAGYNNDVSTLRWITTKIVGH 240
DB 181 LHLPEVLLPCTGHTTSTYLLHGDGRANVYEGDTSAGYNNDVSTLRWITTKIVGH 240
QY 241 PLVIRVRAIGHFVLLTAAPESPMPVYPRSTEVVRSIFGSGSPSLFPSACSTRK 300
DB 241 PLVIRVRAIGHFVLLTAAPESPMPVYPRSTEVVRSIFGSGSPSLFPSACSTRK 300
QY 301 STEHAVPHIMDRMLFGATLDDQAFCCSRMLTYLRGISYKVTGALVANEGMNASDRL 360
DB 301 STEHAVPHIMDRMLFGATLDDQAFCCSRMLTYLRGISYKVTGALVANEGMNASDRL 360
QY 361 TAXITAAVLTICHOYRLRQATISKMRRLGVENHAKFTIRLXSNLPEKSGRYTIGROQ 420
DB 361 TAXITAAVLTICHOYRLRQATISKMRRLGVENHAKFTIRLXSNLPEKSGRYTIGROQ 420
QY 421 FYAOCRRMLASAGFHLDPRLVDESVPCKRTFLKVKAGKCCCFMRLGEOCTCELEPAE 480
DB 421 FYAOCRRMLASAGFHLDPRLVDESVPCKRTFLKVKAGKCCCFMRLGEOCTCELEPAE 480
QY 481 GLVGDHGDNEAYEGSEVDPAEPALHDVSGTYAVHGHOLEALYRALNVPDITAAARSRLT 540
DB 481 GLVGDHGDNEAYEGSEVDPAEPALHDVSGTYAVHGHOLEALYRALNVPDITAAARSRLT 540
QY 541 ATVELVASPDRLECKRTVGNKTEFTTVYDGAHLEANGPROYLSLSDASQSGASHSLT 600
DB 541 ATVELVASPDRLECKRTVGNKTEFTTVYDGAHLEANGPROYLSLSDASQSGASHSLT 600
QY 601 YELTPAGLOVRISNGLDCTATFPFGAAPSAPGEVAAFCSALYRYNRFTHSHSLTGLIM 660
DB 601 YELTPAGLOVRISNGLDCTATFPFGAAPSAPGEVAAFCSALYRYNRFTHSHSLTGLIM 660
QY 661 LHPBGLGIFPPFSGHWTESANPFCGEGTLYTRTWSTSGSSDPSPEAAAPAMAATPG 720
DB 661 LHPBGLGIFPPFSGHWTESANPFCGEGTLYTRTWSTSGSSDPSPEAAAPAMAATPG 720
QY 721 LPHSTPPVSDIHWLPSPSEEPDYDAAPVPAPDPAGLPGPVLLPPLPPPVAKPSIPPS 780
DB 721 LPHSTPPVSDIHWLPSPSEEPDYDAAPVPAPDPAGLPGPVLLPPLPPPVAKPSIPPS 780
QY 781 RNRRLLYTPDGAKYVAGSLFESDCDWLVNANSPGHRPGGGLCHAFYORFPAEPFTEI 840
DB 781 RNRRLLYTPDGAKYVAGSLFESDCDWLVNANSPGHRPGGGLCHAFYORFPAEPFTEI 840
QY 841 MREGIAATTLPRPIIHAAPDYRVEONPKRLAAVRETCSRGTAAVPLLGSGIYQVPV 900

DB 841 MREGIAATTLPRPIIHAAPDYRVEONPKRLAAVRETCSRGTAAVPLLGSGIYQVPV 900
QY 901 SUSPFAMERNHRPGDELYTEPAANMFENAKRAOPVLITTEPTARTANLALIDAATEVG 960
DB 901 SUSPFAMERNHRPGDELYTEPAANMFENAKRAOPVLITTEPTARTANLALIDAATEVG 960
QY 961 RACAGCTISPGIVHYQFTAGVBSGKSRSIQGDVDVVVPTRELNSWRRGFAAFTH 1020
DB 961 RACAGCTISPGIVHYQFTAGVBSGKSRSIQGDVDVVVPTRELNSWRRGFAAFTH 1020
QY 1021 TAARVTIGRRVYIDAPSLPHLLLLHMORASSYHLGDPNQIPALIDEFHAQLVAIRE 1080
DB 1021 TAARVTIGRRVYIDAPSLPHLLLLHMORASSYHLGDPNQIPALIDEFHAQLVAIRE 1080
QY 1081 LAPTSMMXYTHRCPADVCCLIGAYPKIOTTSRVLSLFWNPAPAGOKLVYXQAANKAP 1140
DB 1081 LAPTSMMXYTHRCPADVCCLIGAYPKIOTTSRVLSLFWNPAPAGOKLVYXQAANKAP 1140
QY 1141 GAITVHEAGATFETTTIATADARGLIOSSRAHAIVALTRTEKCVIILADGLLEWGI 1200
DB 1141 GAITVHEAGATFETTTIATADARGLIOSSRAHAIVALTRTEKCVIILADGLLEWGI 1200
QY 1201 SDVIYNNFFLAGGEVXHRPSYIPGNPDQNTGTLQAPPSCQISAYHQLAEELGHRAP 1260
DB 1201 SDVIYNNFFLAGGEVXHRPSYIPGNPDQNTGTLQAPPSCQISAYHQLAEELGHRAP 1260
QY 1261 VAAVLPCPELEOGLLYMQEQLTSDSYLVFELTDIVHCRMAAPSORKAVLSTLVGRYGR 1320
DB 1261 VAAVLPCPELEOGLLYMQEQLTSDSYLVFELTDIVHCRMAAPSORKAVLSTLVGRYGR 1320
QY 1321 RTKLYEAHSDVRESLAREIPTIGPVQATTCLELYELVEMVERGODGSAYLELDCNRDV 1380
DB 1321 RTKLYEAHSDVRESLAREIPTIGPVQATTCLELYELVEMVERGODGSAYLELDCNRDV 1380
QY 1381 SRTTFQKXCKNFTTGETIARHKVQOGISAMKTCALFGPFRALKEKTLALLRPNTFY 1440
DB 1381 SRTTFQKXCKNFTTGETIARHKVQOGISAMKTCALFGPFRALKEKTLALLRPNTFY 1440
QY 1441 GDAYEESYFAAASVAGSGCMVFENDFSEPTQNNPFSGLCEVMBECCMPOMLIRLYHL 1500
DB 1441 GDAYEESYFAAASVAGSGCMVFENDFSEPTQNNPFSGLCEVMBECCMPOMLIRLYHL 1500
QY 1501 VRSAMILQPKESLKGFMKHSGBEGTLLMNTVMMAIIAHCYEERDFRVAAFKGDSDVY 1560
DB 1501 VRSAMILQPKESLKGFMKHSGBEGTLLMNTVMMAIIAHCYEERDFRVAAFKGDSDVY 1560
QY 1561 LCSIDYRQSNMAALLAGCGKLKVDYRPIGLIAGVAVAGLGLDYPVAFGRILSEKMG 1620
DB 1561 LCSIDYRQSNMAALLAGCGKLKVDYRPIGLIAGVAVAGLGLDYPVAFGRILSEKMG 1620
QY 1621 PGPERRAOLRLAVCDFLRLGTINVAQVDSRVYVSGVLVHNLIGMLQTLADGKAHPT 1680
DB 1621 PGPERRAOLRLAVCDFLRLGTINVAQVDSRVYVSGVLVHNLIGMLQTLADGKAHPT 1680
QY 1681 ETIKRVLDTNSIIQROVE 1698
DB 1681 ETIKRVLDTNSIIQROVE 1698

RESULT 2
Q9YK10 PRELIMINARY; PRT; 1708 AA.
AC Q9YK10:
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE NONSTRUCTURAL POLYPROTEIN.
OS Swine hepatitis E virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
OX NCBI_TaxID=63421.
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=MENG;
RA MEDLINE=97420774; PubMed=9275216;
RA Meng X.J., Purcell R.H., Halbur P.G., Lehman J.R., Webb D.M.,
RA Tsareva T.S., Haynes J.S., Thacker B.J., Emerson S.U.;
RT "A novel virus in swine is closely related to the human hepatitis E
RT virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:9860-9865(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MENG;
RX MEDLINE=99030877; PubMed=9811705;
RX Meng X.J., Halbur P.G., Shapiro M.S., Govindarajan S., Bruna J.D.,
RA Mushahar I.K., Purcell R.H., Emerson S.U.;
RT "Genetic and experimental evidence for cross-species infection by
RT swine hepatitis E virus.";
RL J. Virol. 72:9714-9721(1998).
DR EMBL: AF082843; AAC97208.1; -;
DR InterPro: IPR000606; Viral_helicase1.
DR InterPro: IPR002588; Viral_helicase1.
DR InterPro: IPR002589; DUF27.
DR pfam: PF01443; Viral_helicase1; 1.
DR pfam: PF01660; Vmethyltransf; 1.
DR pfam: PF01661; DUF27; 1.
DR SMART: SM00506; A1pp; 1.
KW Polypeptide.
FT CHAIN 55 237 METHYLTRANSFERASE.
FT CHAIN 975 1219 HELICASE.
FT CHAIN 1222 1708 RNA-DIRECTED RNA POLYMERASE.
SQ SEQUENCE 1708 AA; 187406 MW; 7A44E52DCD616130 CRC64;

Query Match 97.8%; Score 8812.5; DB 12; Length 1708;
Best Local Similarity 97.8%; Pred. No. 0;
Matches 1661; Conservative 7; Mismatches 30; Indels 1; Gaps 1;

QY 1 PGTTTAEQALAAANASALANAVVPRFLSRVQTELLINIMQPROLVFREVLMNPIOR 60
DB 10 PGTTTAEQALAAANASALANAVVPRFLSRVQTELLINIMQPROLVFREVLMNPIOR 69
QY 61 VTHNELEYCRARAGCLEYGAHPRISINDPNVLRHCFLEPRVGRDVOQWYSAPTRGPAN 120
DB 70 AHNNELEYCRARAGCLEYGAHPRISINDPNVLRHCFLEPRVGRDVOQWYSAPTRGPAN 129
QY 121 CRRSALRGILPPADRTYCFGFSRCAFAETGVALYSLHDMADVAEAMRHGXTRLYAA 180
DB 130 CRRSALRGILPPADRTYCFGFSRCAFAETGVALYSLHDMADVAEAMRHGXTRLYAA 189
QY 181 LHPPEVLLPPGTYHTTSYLLIHDGRAVYTYEGDSAGYNHDVSLIRAMIRTTKIYGDH 240
DB 190 LHPPEVLLPPGTYHTTSYLLIHDGRAVYTYEGDSAGYNHDVSLIRAMIRTTKIYGDH 249
QY 241 PLVIEVRALIGCHFVLLTAAPESPMPVYPRSTEVVRSIFGPGSFLPSPSASTK 300
DB 250 PLVIEVRALIGCHFVLLTAAPESPMPVYPRSTEVVRSIFGPGSFLPSPSASTK 309
QY 301 SFTHAHPVHNDMLMFGATLDQAFCCSRLMTYLRGISIKVYVGVALVANEGMNASDAL 360
DB 310 SFTHAHPVHNDMLMFGATLDQAFCCSRLMTYLRGISIKVYVGVALVANEGMNASDAL 369
QY 361 TAITTAAYLTICHQRLRTOAISKGRRLGVEHAOKFTTLYSLMPEKSGRDYIPGROLQ 420
DB 370 TAITTAAYLTICHQRLRTOAISKGRRLGVEHAOKFTTLYSLMPEKSGRDYIPGROLQ 429
QY 421 FYAOCRRLMSAGFHLDPRLVLPFDESVPQRCRTFLKRVAGKFCFMRWLGQECTFLERPAE 480
DB 430 FYAOCRRLMSAGFHLDPRLVLPFDESVPQRCRTFLKRVAGKFCFMRWLGQECTFLERPAE 489
QY 481 GLVGDGHNDNEAYEGSEVDPAPBAHLDVSGTYAVHGHQLEFALYRALNVPDIIAARSLT 540
DB 490 GLVGDGHNDNEAYEGSEVDPAPBAHLDVSGTYAVHGHQLEFALYRALNVPDIIAARSLT 549
QY 541 ATVELVASPRLRCRTYLGKTFRTTYVDGAHLEANGPEQYVLSFDSRQSMGAGSHSLT 600

DB 550 ATVELTASPDRLRCRTYLGKTFRTTYVDGAHLEANGPEQYVLSFDSRQSMGAGSHSLT 609
QY 601 YELTPAGLOVRISNSNLDCTATPPPGAPSAABGEVAAPFCASALYRNRFORSLTGLW 660
DB 610 YELTPAGLOVRISNSNLDCTATPPPGAPSAABGEVAAPFCASALYRNRFORSLTGLW 669
QY 661 LMPEGLIGTFPPSPGHIMESANPFCGEGTLYRTSTSGSSDFSPPEAAAPAMATPG 720
DB 670 LMPEGLIGTFPPSPGHIMESANPFCGEGTLYRTSTSGSSDFSPPEAAAPAMATPG 729
QY 721 LPHSTPPVSDIWLPPPESEEOVDAAPVPAPDPAGLPGCVLT -PPPEPVHKPSIPP 779
DB 730 LPHSTPPVSDIWLPPPESEEOVDAAPVPAPDPAGLPGCVLT -PPPEPVHKPSIPP 789
QY 780 SRNRRLTYTPDCAKYAGSLFESDDMLVNASNPNHGRPGGGLCHAFYORFPAPFPTER 839
DB 790 SRNRRLTYTPDCAKYAGSLFESDDMLVNASNPNHGRPGGGLCHAFYORFPAPFPTER 849
QY 840 IMREGLAAYTLTPRPIITHAAPDYRVEQNPRLAAVRETCSSRGTAAYPLGSGIYQV 899
DB 850 IMREGLAAYTLTPRPIITHAAPDYRVEQNPRLAAVRETCSSRGTAAYPLGSGIYQV 909
QY 900 VLSLFDAMERNHRPGDELTYLEPAAMFEANKPAQPALTTTEDTARTANLALETDAATEV 959
DB 910 VLSLFDAMERNHRPGDELTYLEPAAMFEANKPAQPALTTTEDTARTANLALETDAATEV 969
QY 960 GRACAGCTISPGIVHQTFTAGVSGSKRSIQOGDVVVVYPTRELNSRRRGFAFTP 1019
DB 970 GRACAGCTISPGIVHQTFTAGVSGSKRSIQOGDVVVVYPTRELNSRRRGFAFTP 1029
QY 1020 HTAARVYIGRRVYIDEAPSLPPHLLLMORASSVHLGDPNOIPATIDFAGLVPAIRP 1079
DB 1030 HTAARVYIGRRVYIDEAPSLPPHLLLMORASSVHLGDPNOIPATIDFAGLVPAIRP 1089
QY 1080 ELAPTSWMXYTHRCPADVCBLRGAYPKIQTSRVLKSLFWNEPAIGOKLVYQAANKAN 1139
DB 1090 ELAPTSWMXYTHRCPADVCBLRGAYPKIQTSRVLKSLFWNEPAIGOKLVYQAANKAN 1149
QY 1140 PGATVTHEAGATFETTTIATDARGLIOSSRAHALVALTRHTEKVIILDAQGLREVG 1199
DB 1150 PGATVTHEAGATFETTTIATDARGLIOSSRAHALVALTRHTEKVIILDAQGLREVG 1209
QY 1200 ISDVIYNNFVLAGEVGHXRPVYPRGNPDONICTLOAFPSCQISAYHOLAELGRPA 1259
DB 1210 ISDVIYNNFVLAGEVGHXRPVYPRGNPDONICTLOAFPSCQISAYHOLAELGRPA 1269
QY 1260 PVAAYLPPCELEBOGLYMPQELTVSDVLFELTIDVHCRMAAPSORKAVLSTLYGRYG 1319
DB 1270 PVAAYLPPCELEBOGLYMPQELTVSDVLFELTIDVHCRMAAPSORKAVLSTLYGRYG 1329
QY 1320 RRTKLYEAAHSDVRESIARFTPTIGVQATTCCLYELVEAMVEKQDGSVLELDLCNRD 1379
DB 1330 RRTKLYEAAHSDVRESIARFTPTIGVQATTCCLYELVEAMVEKQDGSVLELDLCNRD 1389
QY 1380 VSRITTFQKCKNKFETGETTAHGVGOGISAMSTPCALGPMFRALTEKETLILPPNIF 1439
DB 1390 VSRITTFQKCKNKFETGETTAHGVGOGISAMSTPCALGPMFRALTEKETLILPPNIF 1449
QY 1440 YGDAYEESVFAAASGASCMVEENDFSEFDSTONNNSLGLCYVMECCGMQWILILYH 1499
DB 1450 YGDAYEESVFAAASGASCMVEENDFSEFDSTONNNSLGLCYVMECCGMQWILILYH 1509
QY 1500 LVRSAMITLQAPKESLKGFMKKHSGEPSTLLMNTVMNNAITAHCEYFDFRVAARFGDSDV 1559
DB 1510 LVRSAMITLQAPKESLKGFMKKHSGEPSTLLMNTVMNNAITAHCEYFDFRVAARFGDSDV 1569
QY 1560 VLSGDYRQSRNAALILAGCGKLKLVDRPITGLAGVVAAGLGLTPPVVFRPAGRLSKRW 1619
DB 1570 VLSGDYRQSRNAALILAGCGKLKLVDRPITGLAGVVAAGLGLTPPVVFRPAGRLSKRW 1629
QY 1620 GPGERAEQRLAVCDFRLGTLNVAQCVVSVYVSGPLVNLIGMLQTIADGKAHF 1679
DB 1630 GPGERAEQRLAVCDFRLGTLNVAQCVVSVYVSGPLVNLIGMLQTIADGKAHF 1689

QY 1680 TETIKPVLDTNSIQORVE 1698
1690 TETIKPVLDTNSIIORVE 1708

RESULT 3
QYLR1
ID QYLR1 PRELIMINARY: PRT: 1708 AA.
AC QYLR1:
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE POLYPEPTIDE.
GN ORF1.
OS Hepatitis E virus.
OC Viruses: ssRNA positive-strand viruses, no DNA stage.
OX NCBI_TaxID=12461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HEV-US2;
RX MEDLINE=98178637; PubMed=9519822;
RA Schauder G.G., Dawson G.J., Erker J.C., Kwo P.Y., Knigge M.F.,
RA Smalley D.L., Rosenblatt J.E., Desai S.M., Mushahwar I.K.;
RT "The sequence and phylogenetic analysis of a novel hepatitis E virus
RT isolated from a patient with acute hepatitis reported in the United
RT States.";
RL J. Gen. Virol. 79:0-0(0).
RM [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HEV-US2;
RX MEDLINE=99190429; PubMed=10092008;
RA Erker J.C., Desai S.M., Schauder G.G., Dawson G.J., Mushahwar I.K.;
RT "A hepatitis E virus variant from the United States: molecular
RT characterization and transmission in cynomolgus macaques.";
RL J. Gen. Virol. 80:681-690(1999).
DR EMBL: AF060669; AAD15815.1; -;
DR InterPro: IPR000606; Viral_helicase1.
DR InterPro: IPR002588; V_methyltransf.
DR InterPro: IPR002589; DUF27.
DR Pfam: PF01443; Viral_helicase1; 1.
DR Pfam: PF01660; Methyltransf; 1.
DR Pfam: PF01661; DUF27; 1.
DR SMART: SM00506; A1pp; 1.
KW Polyprotein.
SQ SEQUENCE 1708 AA; 187260 MW; BE0DF5EEEDAB547 CRC64;

Query Match 97.4%; Score 8780.5; DB 12; Length 1708;
Best Local Similarity 97.4%; Pred. No. 0;
Matches 1655; Conservative 8; Mismatches 35; Indels 1; Gaps 1;

QY 1 PGITAIIOALAAANSALANAVVRPFLSRVOTELINLMOPROLYPEPEVLMNHP10R 60
Db 10 PGITAIIOALAAANSALANAVVRPFLSRVOTELINLMOPROLYPEPEVLMNHP10R 69
QY 61 VIHNELEYCARAGRCLEVGANPSINDPNVLRHCEFLRPGRDQVQRYSAPTGPAAN 120
Db 70 VIHNELEYCARAGRCLEVGANPSINDPNVLRHCEFLRPGRDQVQRYSAPTGPAAN 129
QY 121 CRSLALRGLPADRTTYCFDGSRCAPAAETGYALYSLHDLMPADVAEAMARHGTRLYAA 180
Db 130 CRSLALRGLPADRTTYCFDGSRCAPAAETGYALYSLHDLMPADVAEAMARHGTRLYAA 189
QY 181 LHLPEVLLPGETYHTSTSLIHGDRAVVTYEGDTSAGYNVDLSILRAMITTKIYGDH 240
Db 190 LHLPEVLLPGETYHTSTSLIHGDRAVVTYEGDTSAGYNVDLSILRAMITTKIYGDH 249
QY 241 PLVIERVAIGCHFVLLTAAPEPSMPYVPRSTEVYVRSIFPGSGPSLFPACSTK 300
Db 250 PLVIERVAIGCHFVLLTAAPEPSMPYVPRSTEVYVRSIFPGSGPSLFPACSTK 309
QY 301 STEHAVPVHMDRLMFGATLDDQAFCCSRMLMTYLRGISYKVTGALVANEGMNASDAL 360

Db 310 STEHAVPVHMDRLMFGATLDDQAFCCSRMLMTYLRGISYKVTGALVANEGMNASDAL 369
QY 361 TAXITAAVITICHORLRTQALISKGRRLGVEHAQCFITRLYSMLEKSGROYIPROQ 420
Db 370 TAXITAAVITICHORLRTQALISKGRRLGVEHAQCFITRLYSMLEKSGROYIPROQ 429
QY 421 FYAOCRRMLSAGFHLDPRLVFEDESVPCRTFLFKVAKGFCFMMLOECTCFLEPAE 480
Db 430 FYAOCRRMLSAGFHLDPRLVFEDESVPCRTFLFKVAKGFCFMMLOECTCFLEPAE 489
QY 481 GLVGDGHNDNEAYEGSEVDPAEPALHDVSGTYAVHGHOLEALYRALNPQDIAAARSRLT 540
Db 490 GLVGDGHNDNEAYEGSEVDPAEPALHDVSGTYAVHGHOLEALYRALNPQDIAAARSRLT 549
QY 541 ATVEVLAAPDRLECRVTLGNKTFRTYVDGAHLLEANGPEQOYVLSPASQSGAGSHSLT 600
Db 550 ATVEVLAAPDRLECRVTLGNKTFRTYVDGAHLLEANGPEQOYVLSPASQSGAGSHSLT 609
QY 601 YELTPAGLQVRISNGLDCTATFPPGAPSAAPGEVAACFSLYRNFRTOHSLTGLM 660
Db 610 YELTPAGLQVRISNGLDCTATFPPGAPSAAPGEVAACFSLYRNFRTOHSLTGLM 669
QY 661 LHPEGLIGIFPPSPGHIMESANPFCGECTLYTRWSTSGFSDESPPPAAPAMAAATPG 720
Db 670 LHPEGLIGIFPPSPGHIMESANPFCGECTLYTRWSTSGFSDESPPPAAPAMAAATPG 729
QY 721 LPHSTPPVSDIVLPPSPSEFOYDAAPVPADPDAGLPEPVLT -PPPPPVHKSIIPP 779
Db 730 LPPVPPVSDIVLPPSPSESHVDAASVPSPPEGLTSPVLTTPPPPPVPRKAPATSP 789
QY 780 SNRRRLTYTPPGAKYAGSLFESDCDMLVMSNPGHRGGGLCHAFYORPEAPYPTPF 839
Db 790 PSTRRLTYTPPGAKYAGSLFESDCDMLVMSNPGHRGGGLCHAFYORPEAPYPTPF 849
QY 840 IREGLIAATLTPRPIIHAAPDYRVEONPKRLAAAYRETCSRKGTAAYPLLGSIYQVP 899
Db 850 IREGLIAATLTPRPIIHAAPDYRVEONPKRLAAAYRETCSRKGTAAYPLLGSIYQVP 909
QY 900 VLSFDAMERNRHPGDELTLTPAANMFANPAPQVLTITDPTRTANLALIDEADATPV 959
Db 910 VLSFDAMERNRHPGDELTLTPAANMFANPAPQVLTITDPTRTANLALIDEADATPV 969
QY 960 GRACAGCTISPGIVHYQFTAGVPGSGKRSIQGDVDVYVPTRELNSMRRRGFAATP 1019
Db 970 GRACAGCTISPGIVHYQFTAGVPGSGKRSIQGDVDVYVPTRELNSMRRRGFAATP 1029
QY 1020 HTAAVYITGRVYIDEASPLPHLLLMQARSSVHLGDPNOIPALIDEHAGLYPAIRP 1079
Db 1030 HTAAVYITGRVYIDEASPLPHLLLMQARSSVHLGDPNOIPALIDEHAGLYPAIRP 1089
QY 1080 ELAPTSMWVTHRCPADVCELRGAYPKIQTSTRVLSRSLFEMNEPALGOKLYTOAKKAN 1139
Db 1090 ELAPTSMWVTHRCPADVCELRGAYPKIQTSTRVLSRSLFEMNEPALGOKLYTOAKKAN 1149
QY 1140 PGATVHEAGATFTETTLIATADARGLIOSSRAHAIVALTHTKTCVTLDAAGLREVG 1199
Db 1150 PGATVHEAGATFTETTLIATADARGLIOSSRAHAIVALTHTKTCVTLDAAGLREVG 1209
QY 1200 ISDVIVNNEFLAGGEVGAHRBSVIFRGNDONLGLQAFPPSCQISAHQOLAEELGHRPA 1259
Db 1210 ISDVIVNNEFLAGGEVGAHRBSVIFRGNDONLGLQAFPPSCQISAHQOLAEELGHRPA 1269
QY 1260 PYAAVLPQCPPELQGLLWPOELTYSDSVLVELLDIYHCRMAAPSORAVLSTLVGRG 1319
Db 1270 PYAAVLPQCPPELQGLLWPOELTYSDSVLVELLDIYHCRMAAPSORAVLSTLVGRG 1329
QY 1320 RRTKLYEAHSDVRESLARFTPTIGPVQATTCLELVEIYEAWEKQDSAVLELDLCND 1379
Db 1330 RRTKLYEAHSDVRESLARFTPTIGPVQATTCLELVEIYEAWEKQDSAVLELDLCND 1389
QY 1380 VSRITFFQKXCKFTTGETIAHGKVGQGISAMSKTFCALFGWPAIEKEITALLPNTF 1439

Db 1390 VSRITFEQKCNKFTTGETTIAHGKVGQGISAMSKTECALFGPWFRAIEKEILLAPNIF 1449
QY 1440 YGSAVESVFAAASGAGSGMVEENDESESDSONNPSIGLECYVMEECGMQWILRLYH 1499
Db 1450 YGSAVESVFAAASGAGSGMVEENDESESDSONNPSIGLECYVMEECGMQWILRLYH 1509
QY 1500 LVMSAMTLOAPKESLKGFMKHSGEPTLLMNTVMNMAIIAHCEYEFDFVAAAFKGDVS 1559
Db 1510 LVMSAMTLOAPKESLKGFMKHSGEPTLLMNTVMNMAIIAHCEYEFDFVAAAFKGDVS 1569
QY 1560 VLCSDYRQSNMAALAGCGLKLVDRPIGLYAGVVAADGLTLPDVRFAGRLSEKNV 1619
Db 1570 VLCSDYRQSNMAALAGCGLKLVDRPIGLYAGVVAADGLTLPDVRFAGRLSEKNV 1629
QY 1620 GGPPEAEQRLAVCDPLRLTNVAQYCVVSVRYGVSPGLVHNLIGMLQTADGAHF 1679
Db 1630 GGPPEAEQRLAVCDPLRLTNVAQYCVVSVRYGVSPGLVHNLIGMLQTADGAHF 1689
QY 1680 TETIKPVLDTNSIIOHVE 1698
Db 1690 TENIKPVLDTNSIIOHVE 1708
RESULT 4
Q91VZ9 PRELIMINARY; PRT: 1707 AA.
AC Q91VZ9;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE POLYPROTEIN.
OS Hepatitis E virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
ON NCBI_TaxId=12461;
OX [1]
RN RP
RC STRAIN=H1;
RA Harrison T.J.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RN
RC STRAIN=H1;
RX MEDLINE=20318653; PubMed=10859372;
RT "The complete sequence of hepatitis E virus genotype 4 reveals an
RT alternative strategy for translation of open reading frames 2 and 3.";
RL J. Gen. Virol. 81:1675-1686(2000).
DR EMBL: AJ272108; CAB83209.1; -.
DR InterPro: IPR002589; DUF27.
DR InterPro: IPR000606; Viral_helicase.
DR InterPro: IPR002588; V_methyltransf.
DR Pfam: PF01461; DUF27.1.
DR Pfam: PF01443; Viral_helicase.1.
DR Pfam: PF01660; Vmethyltransf.1.
DR SMART: SM00506; A1pp.1.
SO SEQUENCE 1707 AA, 187296 MW, 2A80F303AB96C4EC CRC64;

Query Match 85.8%; Score 7730; DB 12; Length 1707;
Best Local Similarity 83.9%; Pred. No. 0;
Matches 1444; Conservative 91; Mismatches 140; Indels 46; Gaps 5;

QY 1 PGTTTAEQAALAAANSALANAVVRPFLSRVOTELLINIMOPROLVFREPVLNHPRIOR 60
Db 10 PGVTTAEQAALAAANSALANAVVRPFLSRVOTELLINIMOPROLVFREPVLNHPRIOR 69
QY 61 VIHNELEYCARARAGCLLEVGAHPRSTINDPNVYHRCFLRPVGDVQRWTSAPTRGRAN 120
Db 70 VIHNELEYCARARAGCLLEVGAHPRSTINDPNVYHRCFLRPVGDVQRWTSAPTRGRAN 129
QY 121 CRSSALRGPLPADPTTCFDFSGSCAFEAETGVALYSLHDLMPADVAAMARHGXTRLYAA 180
Db 130 CRSSALRGPLPADPTTCFDFSGSCAFEAETGVALYSLHDLMPADVAAMARHGXTRLYAA 189

QY 181 LHLPEVLLPGTYHTTSYLLIHGDRAVVTYEDGTSGAGYNHDSILRAMIRTTKIYGDH 240
Db 190 LHLPEVLLPGTYHTTSYLLIHGDRAVVTYEDGTSGAGYNHDSILRAMIRTTKIYGDH 249
QY 241 PLVIERVRAIGCHEVLLTLTAAPESPMPVYVPRSTEVYVRSIFPGGSPSLFSPASCTK 300
Db 250 PLVIERVRAIGCHEVLLTLTAAPESPMPVYVPRSTEVYVRSIFPGGSPSLFSPASCTK 309
QY 301 STEHAVVHITWDRMLFGLALDDOAFCCSRMLTYLRKISIKYVYGVALVANEAGNASDAL 360
Db 310 STEHAVVHITWDRMLFGLALDDOAFCCSRMLTYLRKISIKYVYGVALVANEAGNASDAL 369
QY 361 TAYTAAVLTTCCHORYLRTQALISKGMRLCYLEHAKOITLYLSMLFEKSGSDYIPGROLO 420
Db 370 TAYTAAVLTTCCHORYLRTQALISKGMRLCYLEHAKOITLYLSMLFEKSGSDYIPGROLO 429
QY 421 FYAQCRRLWSAGFHLDPRLVLFDESVPCRCRTFLKKAAGFCCPMRLIGOECTCFLEPAE 480
Db 430 FYAQCRRLWSAGFHLDPRLVLFDESVPCRCRTFLKKAAGFCCPMRLIGOECTCFLEPAE 489
QY 481 GLVGDHGHNDAYEGSEVDPAEPALHDVSGTYAVHGHOLEALYRALVPODIAARASRLT 540
Db 490 GRVGEQGYDNDAEFGSDIDPAEFAVSIAGSYIVTGSQLOPLYOALGIPSDLAARASRLT 549
QY 541 ATVELVASPDLRCRTYVLGKTFRTTYVDGAHLEANGPEQYVLSFSDASROSGASHLT 600
Db 550 ATVELVASPDLRCRTYVLGKTFRTTYVDGAHLEANGPEQYVLSFSDASROSGASHLT 609
QY 601 YELTPAGLQYRISNGLDCTATFPFGAPASAPAEVAFSCALRYNRFTQRLHSLTGLW 660
Db 610 YTLTSLAGLEVHVSAGLDCCKVYRQSGVAPASAGEVAFSCALRYNRFTQRLHSLTGLW 669
QY 661 LHPGILLGIFPPESPGLHWSANPFCGEGTLYRTWSTSGFSSDFSPPEAAAPMAATPG 720
Db 670 YHPGILLGIFPPESPGLHWSANPFCGEGTLYRTWSTSGFSSDFSPPEAAAPMAATPG 724
QY 721 LPHSTPVSIDVLPPESE---EQVDAAP-----VPPAPDPAGLP----- 758
Db 725 LPHSTPVSIDVLPPESE---EQVDAAP-----VPPAPDPAGLP----- 770
QY 759 -GPVLTFRPPRPVYKPSIIPPSRNRRLTYPPGAKYVAGSLFESDCDMLVNASNPGHR 817
Db 771 SGAPIAPPAALPYTHLSGPR---RRLLHYTPGSKVYAGSLFESCTWLVMNASNPGHR 826
QY 818 PGGLCHAFYQRFPEAEYPTFEIMREGIAAYTLPRPIHAHVAADYREONPKRLAEAYR 877
Db 827 PGGLCHAFYQRFPEAEYPTFEIMREGIAAYTLPRPIHAHVAADYREONPKRLAEAYR 886
QY 878 ETCSRRTAAYPLDLSGTYOYVPSLSFDAMERNRPGDELTLTPPAANWFPAKPAQVYL 937
Db 887 ETCSRRTAAYPLDLSGTYOYVPSLSFDAMERNRPGDELTLTPPAANWFPAKPAQVYL 946
QY 938 TITEDTARTANLALDEIDAATFVGACAGCTSPGIVHVOFTAGVPGSGSKSIQOGPDVY 997
Db 947 TITEDTARTANLALDEIDAATFVGACAGCTSPGIVHVOFTAGVPGSGSKSIQOGPDVY 1006
QY 998 VVPTRELNSWRRRGFAAFTPHTAARYTIGRRVVIDEAPSLPHLLLLHMQRASSVHL 1057
Db 1007 IVVPTRELNSWRRRGFAAFTPHTAARYTIGRRVVIDEAPSLPHLLLLHMQRASSVHL 1066
QY 1058 GDPNQTALIDFENAGLVPAIRBELAPTSWXYVTHRCPADVCCELLRGAPKQTTSRYLRS 1117
Db 1067 GDPNQTALIDFENAGLVPAIRBELAPTSWXYVTHRCPADVCCELLRGAPKQTTSRYLRS 1126
QY 1118 LFWNEPAIGOKLVYTOAKANPAGAITVHEOGATFTETTITATADAGLQSSRAHIV 1177
Db 1127 LFWNEPAIGOKLVYTOAKANPAGAITVHEOGATFTETTITATADAGLQSSRAHIV 1186
QY 1178 ALTRHTEKCVTLDAAGLREVGISDVIVNNEFLAGVEGXHRSVIRPGNDONLIGTLQA 1237
Db 1187 ALTRHTEKCVTLDAAGLREVGISDVIVNNEFLAGVEGXHRSVIRPGNDONLIGTLQA 1246

QY 1238 FPPSCQISAVHQAELGHRPAVAALPPCELEGGLLYMPQELTVSDSVLVEELTDIV 1297
DB 1247 FPPSCQISAVHQAELGHRPAVAALPPCELEGGLLYMPQELTVSDSVLVEELTDIV 1306
QY 1298 HCMAPPSQKAVSLTVGNGYGRRTKLYEAHSDVRESLARFTPTGPGVATTCCELYEY 1357
DB 1307 HCMAPPSQKAVSLTVGNGYGRRTKLYEAHSDVRESLARFTPTGPGVATTCCELYEY 1366
QY 1358 EAMVEKGODSAYLELDLCNRDVSRTTFQKCNKFTTGETTIAHGKVGQISAMSKTFC 1417
DB 1367 EAMVEKGODSAYLELDLCNRDVSRTTFQKCNKFTTGETTIAHGKVGQISAMSKTFC 1426
QY 1418 LKSPMRALKEKELLALPPNITFYGDAYEESVFAAVSAGSCWVFENDESEEDSTONNES 1477
DB 1427 LKSPMRALKEKELLALPPNITFYGDAYEESVFAAVSAGSCWVFENDESEEDSTONNES 1486
QY 1478 LGLECYVMECCGPMOLIRLYHLYRSAMLLQAKESLKGPMKHSDEPGILLMNTYWNNA 1537
DB 1487 LGLECYVMECCGPMOLIRLYHLYRSAMLLQAKESLKGPMKHSDEPGILLMNTYWNNA 1546
QY 1538 IIAHCEYFDFRYAAEFKGDSDSVLCSQDYRSRMAALLIAGCGILKLVYKPIGLYAGVYV 1597
DB 1547 VIAHCYEFBDLKYAAFKGDSVVLCSQDYRSRMAALLIAGCGILKLVYKPIGLYAGVYV 1606
QY 1598 AAGGLTLPVYVFRAGLSEKNMGCPGPRABQLRLAYCDLRLGTLNVAQVCVDVYSRYGV 1657
DB 1607 AAGGLTLPVYVFRAGLSEKNMGCPGPRABQLRLAYCDLRLGTLNVAQVCVDVYSRYGV 1666
QY 1658 SPGLVNLIGMLQTLADGKAHFTETIKPVLDTLNSITIORE 1698
DB 1667 SPGLVNLIGMLQTLADGKAHFTETIKPVLDTLNSITIORE 1707

RESULT 5
QY 09E8G6 PRELIMINARY: PRT: 1685 AA.
AC 09E8G6:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NONSTRUCTURAL POLYPROTEIN (FRAGMENT).
OS Hepatitis E virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
OX NCBI_Taxid=12461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AB-28;
RA van Cuyck-Gandre H., Zhang H.Y., Tsarev S.A., Warren R.L.,
RA Caudill J.D., Snellings N.D., Begot L., Innis B.L., Longet C.F.,
RA Tichenhurst J.;
RT "Phylogenetically distinct hepatitis E viruses (HEC) in Pakistan";
RL Am. J. Trop. Med. Hyg. 0:0-0(1999).
DR EMBL: AF158822; AAG16764.1;
DR InterPro: IPR002589; DUF27.
DR InterPro: IPR000606; Viral_helicase1.
DR InterPro: IPR002588; V_methyltransf.
DR Pfam: PF01443; Viral_helicase1; 1.
DR Pfam: PF01660; Vmethyltransf; 1.
DR SMART: SM00506; Alpp; 1.
KM Polyprotein.
FT NON_TER 1
SQ SEQUENCE 1685 AA; 184081 MW; 87CA74659104538B CRC64;

Query Match 82.9%; Score 7475; DB 12; Length 1685;
Best Local Similarity 81.6%; Pred. No. 0;
Matches 1400; Conservative 112; Mismatches 155; Indels 48; Gaps 6;

QY 1 PGITTAIDQALAAANSALANAVVVRPFLSRVOTELLINLMQPROLVPRPELVAMHPIDR 60
DB 2 PGITTAIDQALAAANSALANAVVVRPFLSRVOTELLINLMQPROLVPRPELVAMHPIDR 61

QY 61 VIHNELOYCARAGACLEVGAPRSINDPNVLRHCFLPVRGDYQRMYSAPTRGPAN 120
DB 62 VIHNELOYCARAGACLEVGAPRSINDPNVLRHCFLPVRGDYQRMYSAPTRGPAN 121
QY 121 CRSSALRGLEPADRTYCFDGFSCAPAEETGVALYSLHDLMPADYEAAMRHGXLYAA 180
DB 122 CRSSALRGLEPADRTYCFDGFSCAPAEETGVALYSLHDLMPADYEAAMRHGXLYAA 181
QY 181 LHPPEVLLPQGYHHTSYLLIHGDRAVYTYGDPISAGNHDVSLIRAMIRTKLVGDH 240
DB 182 LHPPEVLLPQGYHHTSYLLIHGDRAVYTYGDPISAGNHDVSLIRAMIRTKLVGDH 241
QY 241 PLVIERVRAIGHFVLLTLAAPPSPMPYVPRSTEVYVRSIFGSGSPLEPSSACSTK 300
DB 242 PLVIERVRAIGHFVLLTLAAPPSPMPYVPRSTEVYVRSIFGSGSPLEPSSACSTK 301
QY 301 STFHAVPVIHWRMLFEGATLDDQACCSRLMTYLGISTKYVGLVANEENASSEDAL 360
DB 302 STFHAVPVIHWRMLFEGATLDDQACCSRLMTYLGISTKYVGLVANEENASSEDAL 361
QY 361 TAXITAAVLTICHORVLTQATSKGMRRLGVEHAQKFTRLYSMLPEKSGRDYIPGRLO 420
DB 362 TAXITAAVLTICHORVLTQATSKGMRRLGVEHAQKFTRLYSMLPEKSGRDYIPGRLO 421
QY 421 FTAQCRRLWSAGFHLDPRLVTFDESVCRCRTFLKKVAKFCFCFMRILQOECTCLEPAE 480
DB 422 FTAQCRRLWSAGFHLDPRLVTFDESVCRCRTFLKKVAKFCFCFMRILQOECTCLEPAE 481
QY 481 GYVGDCGHNEVESEVDPAEPANLDVGTAAVHGHOLEALYRALNPODIAAARSRLT 540
DB 482 GYVGDCGHNEVESEVDPAEPANLDVGTAAVHGHOLEALYRALNPODIAAARSRLT 541
QY 541 ATVEELVSPDLRECFRTVLNKTFTTYVVDGALHEANGPEQYVLEFSDASQSMGASHLT 600
DB 542 ATVEELVSPDLRECFRTVLNKTFTTYVVDGALHEANGPEQYVLEFSDASQSMGASHLT 601
QY 601 YELTPAGLOVRISNGLDCTATPPPGAGAPSAPEVAACALYRYNRFTQRSLTGLM 660
DB 602 YELTPAGLOVRISNGLDCTATPPPGAGAPSAPEVAACALYRYNRFTQRSLTGLM 661
QY 661 LHPPEGLIGFPPSPGHIMESANPCGECTLVTRIMS-TSGRSSDPS-PEAANA 714
DB 662 LHPPEGLIGFPPSPGHIMESANPCGECTLVTRIMS-TSGRSSDPS-PEAANA 715
QY 715 MATPGLPSTPPVSDIWLPPPESEFOYDAVPP-PADPAGLPGPVLLTPPPPPVAK 773
DB 722 MATPGLPSTPPVSDIWLPPPESEFOYDAVPP-PADPAGLPGPVLLTPPPPPVAK 774
QY 774 PS-----IPPSNRRLLYTYPPDGAKYVAGSLFESDCDWLVNASNPGHRPGGGLC 823
DB 751 PS-----IPPSNRRLLYTYPPDGAKYVAGSLFESDCDWLVNASNPGHRPGGGLC 824
QY 824 HAFYQRFPAFTPTFIMEGLAATLTPRPIIHAVAPRYRYEONPKRLAAYRRECSRR 883
DB 811 HAFYQRFPAFTPTFIMEGLAATLTPRPIIHAVAPRYRYEONPKRLAAYRRECSRR 884
QY 884 GTRAAVPLGSGITGYOVPSLPPDAMERNHRPGGEIVTEPAAAMPFANKPAQOVLLTTEBT 943
DB 871 GTRAAVPLGSGITGYOVPSLPPDAMERNHRPGGEIVTEPAAAMPFANKPAQOVLLTTEBT 930
QY 944 ARTANLALIDATEVGRACAGCTTSPGIVHVOFTAGVYSGSKSRISIOGDVYVVVPTR 1003
DB 931 ARTANLALIDATEVGRACAGCTTSPGIVHVOFTAGVYSGSKSRISIOGDVYVVVPTR 990
QY 1004 ELRNSWRRRGFAFTPHTAARVYIDEPAPLPHLLILHMQRASSVHLGDPPNOI 1063
DB 991 ELRNSWRRRGFAFTPHTAARVYIDEPAPLPHLLILHMQRASSVHLGDPPNOI 1050
QY 1064 PAIDFEHAGVLAIRPELAPTSWMAVYTHRCPADVCELLRGAPKPIQTSRYVRLSFLWNP 1123
DB 1051 PAIDFEHAGVLAIRPELAPTSWMAVYTHRCPADVCELLRGAPKPIQTSRYVRLSFLWNP 1110
QY 1124 AIGKLVYTOAKKANPGAITVHEAGATFTETIATADAGLTISSRAHIVALTTRHT 1183

```

|||||
Db 1111 AVGQKLVFQAAKAAANGSVTVEAGATYETTTIATADARGLIISRHAHVALJTRHT 1170
QY 1184 EKCVIIDAPGLREVGISDVYNNFLAGENGXHRPSYIPKGNPNQIQTLOAPPPSCQ 1243
Db 1171 EKCVIIDAPGLREVGISDAIYNNFLAGENGXHRPSYIPKGNPNANDTLAAPPSCQ 1230
QY 1244 ISAYHOLAEELGHRPAVAALVPCPELBOGLIYMPOLVTSVYVLELTDIVHCRMAA 1303
Db 1231 ISAYHOLAEELGHRPAVAALVPCPELBOGLIYLPOLVTCDSVVFELTDIVHCRMAA 1290
QY 1304 PSQRKAVLSTLVGRYGRRTKLYEAASDVRESIARFIPTIGVQATTCELVELVEAMVEK 1363
Db 1291 PSQRKAVLSTLVGRYGRRTKLYEAASDVRESIARFIPTIGVQATTCELVELVEAMVEK 1350
QY 1364 GQDSAVLELDLCNRDVSRITTFQKXCNKFTTGTAHGVGQGISAMSKTFCALFGPMF 1423
Db 1351 GQDSAVLELDLCNRDVSRITTFQKXCNKFTTGTAHGVGQGISAMSKTFCALFGPMF 1410
QY 1424 RALEKILALPPNITFYGAYESVAAVSAGSGCMVFENDSESDSTONNLSGLECY 1483
Db 1411 RALEKILALPPNITFYGAYESVAAVSAGSGCMVFENDSESDSTONNLSGLECY 1470
QY 1484 VMEECAPQMWLRILYHLYVSAMWILQAPKESLKGFMKHSCEPCTLLMNTVMNAATIAHCY 1543
Db 1471 VMEECAPQMWLRILYHLYVSAMWILQAPKESLKGFMKHSCEPCTLLMNTVMNAATIAHCY 1530
QY 1544 EFDFEFAVAFKGGDSVYLCSIDYRQSRNAALLAGCGIKLVDRPIGLYAGVVAAPGLGT 1603
Db 1531 EFDFEFAVAFKGGDSIVLCSEYRQSPGAALVILGCGIKLVDRPIGLYAGVVAAPGLGT 1590
QY 1604 LPDVVFAARLSEKNNGPPEPERAEQRLAVCDPLRGLTNAQVCVNVVSAVYVSGVLVH 1663
Db 1591 LPDVVFAARLSEKNNGPPEPERAEQRLAVCDPLRGLTNAQVCVNVVSAVYVSGVLVH 1650
QY 1664 NLIGMLOTIADGKAHFTETIKPVLDTNLSIIOHVE 1698
Db 1651 NLIGMLOTIADGKAHFTESKVPVLDTNLSIIOHVE 1685

RESULT 6
ID 069410 PRELIMINARY: PRT: 1693 AA.
AC 069410:
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE METHYL TRANSFERASE.
OS Hepatitis E virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
OX NCBI_TaxID=12461;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=HEV037;
RA Donati M.C., Fagan E.A., Harrison T.J.;
  Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: X98292; CA66936.1; -
DR InterPro: IPR000606; Viral_helicase1.
DR InterPro: IPR002589; V_methyltransf.
DR pfam: PF01443; Viral_helicase1. 1.
DR pfam: PF01660; Vmethyltransf. 1.
DR pfam: PF01661; DUF27. 1.
DR SMART: SM00506; Alip. 1.
DR transbase.
SQ SEQUENCE 1693 AA; 185190 MW; C347C2436DED9B23 CRC64;

Query Match      82.9%; Score 7468; DB 12; Length 1693;
Best Local Similarity 82.3%; Pred. No. 0;
Matches 1402; Conservative 109; Mismatches 167; Indels 26; Gaps 7;

QY 1 PGITTAIEQAALAAANSALANAVVVRPFLSRVQTEILINLMQPRQLVFRREVLMNHP IQR 60
```

```

|||||
Db 10 PGITTAIEQAALAAANSALANAVVVRPFLSRVQTEILINLMQPRQLVFRREVLMNHP IQR 69
QY 61 VHNELBOYCRARAGCLEYGAHPRSTINDPNVLRHCFLRPVGRDQWRYSAPTRGPAAN 120
Db 70 VHNELBOYCRARAGCLEYGAHPRSTINDPNVLRHCFLRPVGRDQWRYSAPTRGPAAN 129
QY 121 CRSSALRGILPPARRTYCFQSEFSCAPAEETGVALYSLHDIIMPADVAAMRHGXTRLYAA 180
Db 130 CRSSALRGILSAARTYCFQSEFSCAPAEETGVALYSLHDIIMPADVAAMRHGXTRLYAA 189
QY 181 LHPPEVLPRPTGYHTTSYLLIDGDAVYTYEGDSAGYNHDSILIRAITTKIYGDH 240
Db 190 LHPPEVLPRPTGYHTTSYLLIDGDAVYTYEGDSAGYNHDSILIRAITTKIYGDH 249
QY 241 PLVIERVRAIGCHFVLLTAAPESPMPVYPRSTEVYRSIFPGSGSPLEPPSACSTK 300
Db 250 PLVIERVRAIGCHFVLLTAAPESPMPVYPRSTEVYRSIFPGSGSPLEPPSACSTK 309
QY 301 STTHAIVPHIWDRLMFLGATLDQAFCCSRMLTYLRGISKYVYGVALYANEGNNSDAL 360
Db 310 STTHAIVPHIWDRLMFLGATLDQAFCCSRMLTYLRGISKYVYGVALYANEGNNSDAL 369
QY 361 TAXITTAAYLTICHOBYLRTQAIKGMRLGVEHAKOFTFLYLSWLFPEKSGDYIPGRQL 420
Db 370 TAXITTAAYLTICHOBYLRTQAIKGMRLGVEHAKOFTFLYLSWLFPEKSGDYIPGRQL 429
QY 421 FYAOCRRMLISAGFHLDPRVLVDFESVPCRCRTFLKRYVAFCCFMRVLGQECTFLEPAE 480
Db 430 FYAOCRRMLISAGFHLDPRVLVDFESVPCRCRTFLKRYVAFCCFMRVLGQECTFLEPAE 489
QY 481 GLVGDHCHDNEATBEGSVDPDAEPRAHLDVSTYAVHGQLEALYLRALNVPQDIAARSRLT 540
Db 490 GLVGDHCHDNEATBEGSVDPDAEPRAHLDVSTYAVHGQLEALYLRALNVPQDIAARSRLT 549
QY 541 ATVELVASPRLCRVYLGKRTFTTYVVDGAHLEANGPEQYVLSFDSRQSMGASHLT 600
Db 550 ATVELVASPRLCRVYLGKRTFTTYVVDGAHLEANGPEQYVLSFDSRQSMGASHLT 609
QY 601 YELTPAGLOVRISNGIDCTATPEPPGAPSAADGEVAFCALYRNYRTQSRSLTGLW 660
Db 610 YELTPAGLOVRISNGIDCTATPEPPGAPSAADGEVAFCALYRNYRTQSRSLTGLW 669
QY 661 LHPPEGLITPRPSGHIWISANPFCGEGTLYRTWS-TSGFSSDFP-----PEAAPA 714
Db 670 LHPPEGLITPRPSGHIWISANPFCGEGTLYRTWS-TSGFSSDFP-----PEAAPA 729
QY 715 MAATPGLPHSTPPVSDIWMVLPPESEEFQVDAAPVPAPADPAGLDPGPVLTLPPEPPVHKR 774
Db 730 MAATPGLPHSTPPVSDIWMVLPPESEEFQVDAAPVPAPADPAGLDPGPVLTLPPEPPVHKR 773
QY 775 SIPEPSRNRRLTYTPDGAKVYAGSLFESDCDMLVNASNGHPRGGGLCHAFYQRPFAE 834
Db 774 SIPEPSRNRRLTYTPDGAKVYAGSLFESDCDMLVNASNGHPRGGGLCHAFYQRPFAE 829
QY 835 YPTEFLIREGLAAYTLTPRTITAAVADRYVEQNPRLKLAAYRTCCRGTAAYPLIGSG 894
Db 830 YPTEFLIREGLAAYTLTPRTITAAVADRYVEQNPRLKLAAYRTCCRGTAAYPLIGSG 889
QY 895 IYOVVLSFSDAMERNRNPDELYLTTEPAANMFPAAPVLTITDTRTATYALEID 954
Db 890 IYOVVLSFSDAMERNRNPDELYLTTEPAANMFPAAPVLTITDTRTATYALEID 949
QY 955 AATEVGRACAGCTISPGIVHYQTAGVYSGSKSRISIQGDVVVVVYPTRELNSWRRRGF 1014
Db 950 AATEVGRACAGCTISPGIVHYQTAGVYSGSKSRISIQGDVVVVVYPTRELNSWRRRGF 1009
QY 1015 AATFPHTAARVITIGRRVVIDEAPSLPRHLLLHMRKSSVHLLGDPNQIATIDEPHAGLV 1074
Db 1010 AATFPHTAARVITIGRRVVIDEAPSLPRHLLLHMRKSSVHLLGDPNQIATIDEPHAGLV 1069
QY 1075 PATRPELATPSSMXVTHRCADVCCELIRGAYPKIQTSRYLSLFWNEPAGIGOKLYVTOA 1134
Db 1070 PATRPELATPSSMXVTHRCADVCCELIRGAYPKIQTSRYLSLFWNEPAGIGOKLYVTOA 1124
```

```
Db 1070 PAIRPDLAPTSMWHVTHRCRADYCELLRGAVPMIOITTSRVLSLNGEBAVGOIKLVETQA 1129
Qy 1135 AKAANGATVTHEAGATFETTTIATADARGLIOSSRAHAVALRHTKECVILDPAGL 1194
Db 1130 AKAANGSVTHAGCATTEFTTIIATADARGLIOSSRAHAVALRHTKECVIIDAPGL 1189
Qy 1195 LKEVGISDIYVNNFVLAGEVGHKHPSTVPRGNPDONLGLTLOAFPSCISAYHQLAEEL 1254
Db 1190 LKEVGISDAIVNNFVLAGEIGIHPSTVPRGNPDONLGLTLOAFPSCISAYHQLAEEL 1249
Qy 1255 GHRPAPVAANLPPCPLEGLLGMPELVSDSVLVEFLTDIVHCHMAAPSQKAVLSTL 1314
Db 1250 GHRPAPVAANLPPCPLEGLLGMPELVSDSVLVEFLTDIVHCHMAAPSQKAVLSTL 1309
Qy 1315 VGRYGRRTLYEAHSDVSESLARFPTIGPYOATTCELYELVEAVNEKGGDSAVLELD 1374
Db 1310 VGRYGRRTLYEAHSDVSDVSLARFPTIGPYOATTCELYELVEAVNEKGGDSAVLELD 1369
Qy 1375 LCNRDVSRITFFOKCKNKTGTGTTAHHGVGOGISAMSTFCALFPGMFRALIEKELALL 1434
Db 1370 LCNRDVSRITFFOKCKNKTGTGTTAHHGVGOGISAMSTFCALFPGMFRALIEKELALL 1429
Qy 1435 PNIFGDVDESVFPAASVAGSGVFEENDESEPDSTONNLSLGEVYMECCGMPQL 1494
Db 1430 PGGVETGDAFDVTSAAVAAKASVFEENDESEPDSTONNLSLGECAIMECCGMPQL 1489
Qy 1495 IRLYHLVRSAMTLOAPKESLKGFWKHSGEPTLLMNTVWNNALIAHCYEFDRFAAAR 1554
Db 1490 IRLYHLVRSAMTLOAPKESLKGFWKHSGEPTLLMNTVWNNALIAHCYEFDRFAAAR 1549
Qy 1555 GDSVYVLCSDYRQSRNAALINGCGILKVDYRPIGLYAGVVAAGLGLTPDVVRFAGRL 1614
Db 1550 GDSVYVLCSEYRQSGAAALINGCGILKVDYRPIGLYAGVVAAGLGLTPDVVRFAGRL 1609
Qy 1615 SEKNNGPGERAEQRLAVCDFLRLGTNTVAOYCVDVSVSVSGPLVNLNIGMOTIAD 1674
Db 1610 SEKNNGPGERAEQRLAVCDFLRLGTNTVAOYCVDVSVSVSGPLVNLNIGMOTIAD 1669
Qy 1675 GKAFHTETIKPVLDTNLSIQHVE 1698
Db 1670 GKAFHTESKPYLDLTNLSICRAVE 1693

RESULT 7
ID 09MIL5 PRELIMINARY: PRT: 1693 AA.
AC 09MIL5:
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE NONSTRUCTURAL POLYPROTEIN.
OC Hepatitis E virus.
OS Viruses: ssRNA positive-strand viruses, no DNA stage.
OX NCBI_TaxID=12461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TK15/92;
RX MEDLINE=99049628; PubMed=9833882;
RA Gouvea V., Snellings N., Popok M.J., Longer C.F., Innis B.L.;
RT "Hepatitis E virus: complete genome sequence and phylogenetic analysis
of a Nepali isolate."
RL Virus Res. 57:21-26(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-TK15/92;
RA Gouvea V.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR InterPro: IPR000606; Viral_helicase1.
DR InterPro: IPR002588; V_methyltransf.
DR InterPro: IPR002589; DUF27.
DR Pfam: PF01443; Viral_helicase1.
DR Pfam: PF01660; Vmethyltransf.1.
```

```
DR Pfam: PF01661; DUF27.1.
DR SMART: SM00506; A1pp; 1.
DR PolyProtein.
SQ SEQUENCE 1693 AA; 185349 MW; A895ACD1DAE2FBBD CRC64;

Query Match 82.7%; Score 7457; DB 12; Length 1693;
Best Local Similarity 81.7%; Pred. No. 0;
Matches 1397; Conservative 110; Mismatches 166; Indels 36; Gaps 6;

Qy 1 PGTTTAIEQALAAANSAANAVVPRPSRQTELLINLMQPROLYVFEVLMNPIOR 60
Db 10 PGTTTAIEQALAAANSAANAVVPRPSRQTELLINLMQPROLYVFEVLMNPIOR 69
Qy 61 VHNLEEOYCRARAGCIEVGAHPRSINDPNVLRFCFLPRGRDQRYMSAPTRGPAAN 120
Db 70 VHNLEELCYCRARSGCIEIGAHPRSINDPNVLRHCFILPRGRDQRYMTATPRGPAAN 129
Qy 121 CRRSALRGLPPADRTYCFDGFSCAFPAETGALYSLHDLMPADVAEMARRHGXRLVAA 180
Db 130 CRRSALRGLPPADRTYCFDGFSGCFSPPETGALYSLHDMSPSDVAEMARRHGXRLVAA 189
Qy 181 LHLPEVLLPFGTYHTSTYLLIHDGDAVYVTEGDTSAQYNHDVSLTAMWITKTIVGH 240
Db 190 LHLPEVLLPFGTYHTSTYLLIHDGDAVYVTEGDTSAQYNHDVSLTAMWITKTIVGH 249
Qy 241 PLVIERVRAIGCFHVLTLAARPESPMPYVPRSTREYVRSIFGSGSPSLFPPSACSTR 300
Db 250 PLVIERVRAIGCFHVLTLAARPESPMPYVPRSTREYVRSIFGSGSPSLFPPSACSTR 309
Qy 301 STEHAVVHIWRLMLFGATLDDQAFCCSRMTYLRGISYKTVGALVANEGMNSSEDAL 360
Db 310 STEHAVVHIWRLMLFGATLDDQAFCCSRMTYLRGISYKTVGALVANEGMNSSEDAL 369
Qy 361 TXXITRAAYLTICHQYRLTQALSKGRRLGVENHAKFTIRLXSNLFEKSGRYITGRQO 420
Db 370 TXXITRAAYLTICHQYRLTQALSKGRRLGVENHAKFTIRLXSNLFEKSGRYITGRQO 429
Qy 421 FYAOCRRMLSAGFHLDPRLVLFDESVCRCRFLKLVKVAFCFCFPMRLGDECTCFLEPAE 480
Db 430 FYAOCRRMLSAGFHLDPRLVLFDESVCRCRFLKLVKVAFCFCFPMRLGDECTCFLEPAE 489
Qy 481 GLVGDHNDNEAYEGSEVDPAEPAHLDVSGTYAVHGHOLEALYRALNVPDIIAARSLTL 540
Db 490 GLVGDHNDNEAYEGSEVDPAEPAHLDVSGTYAVHGHOLEALYRALNVPDIIAARSLTL 549
Qy 541 ATVELVASPDLECRVTGKNTFTTYVDGAHLEANGPQOYVLSLSDASQSGASHSLT 600
Db 550 ATVELVASPDLECRVTGKNTFTTYVDGAHLEANGPQOYVLSLSDASQSGASHSLT 609
Qy 601 YELTPAGLOVRISNGLDCTATFPPGGAAPSAPRGVFAAFCSALYVRNRTORHSLTGLM 660
Db 610 YELTPAGLOVRISNGLDCTATFPPGGAAPSAPRGVFAAFCSALYVRNRTORHSLTGLM 669
Qy 661 LHPBGLGLFPPSPGHIWESANPFCGEGTLYTRTWS-TSGFSDSPS-----PRAAPAA 714
Db 670 LHPBGLGLFPPSPGHIWESANPFCGEGTLYTRTWSIDAYSSPARPOLGVSPTSIPS 729
Qy 715 MAATGCLPHSTIPVSDIWLPPPSSEFOVDAAPV-PPAPDPAGLGCPVYVLTTPPPPVNHK 773
Db 720 MAATGCLPHSTIPVSDIWLPPPSSEFOVDAAPV-PPAPDPAGLGCPVYVLTTPPPPVNHK 774
Qy 774 PSIP-----PPSRNRLTLTYPDGAVVYAGSLFESCDMLVYVNSNGHPRGGGLCAHAFOR 829
Db 765 AGAPAITHOMAHRRLLFTYPDGSVVFAGSLFESTCTYMLVYVNSVNDHRRGGGLCAHAFOR 824
Qy 830 FPEAFYPTFEIIRBGLIAYVTLTPRPIIAVAAPDVYEQNPCKLEAAYRETCSRGTAAVP 889
Db 825 YPASDASAFVWRCDAAYVTLTPRPIIAVAAPDVYEQNPCKLEAAYRETCSRGTAAVP 884
Qy 890 LLGSGIYQVPSLFDAMERNHRPGDELTYLTPAANWFEANPAPVLTITEDTARTANL 949
Db 885 LLGSGIYQVPSLFDAMERNHRPGDELTYLTPAANWFEANPAPVLTITEDTARTANL 944
```

```
QY 950 ALEIDATEYGRACACCTISPGIVHQTACVPGSGSKSISIOGDDVYVVPRELNSM 1009
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 945 AIELDADTDVGRACACGCRYPGVYQYQTAGVSGSKSRISITADADVVPVPRELNSM 1004
QY 1010 RRGFAFPHTARVITGRVVIDEAPSLPPLLLHHORASSVHLGGPNIPIDPE 1069
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1005 RRRGFAFPHTARVITGRVVIDEAPSLPPLLLHHORASSVHLGGPNIPIDPE 1064
QY 1070 HAGLVAIPELAPTSMMVYTHRCPADVCELKAGAPKIQTSRVLSLFWNPATGOKL 1129
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1065 HAGLVAIPELAPTSMMVYTHRCPADVCELKAGAPKIQTSRVLSLFWNPATGOKL 1124
QY 1130 VXTQAAKAMPGATVYHEAGATFETTTIATDANGLIOSRAHAIVALTHTTEKCVIL 1189
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1125 VFTQAKKAMPGATVYHEAGATFETTTIATDANGLIOSRAHAIVALTHTTEKCVIL 1184
QY 1190 DAGGLREVGISDVIVNNFELAGGEVGHRSVTPRGNPONLGTLOAPPSCQISAYHQ 1249
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1185 DAGGLREVGISDVIVNNFELAGGEVGHRSVTPRGNPONLGTLOAPPSCQISAYHQ 1244
QY 1250 LAEELGRAPVAVAVLPCPELEEGLLYMPQELVSDSVLVEFELTDIVHCRMAAPSGRKA 1309
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1245 LAEELGRAPVAVAVLPCPELEEGLLYMPQELVSDSVLVEFELTDIVHCRMAAPSGRKA 1304
QY 1310 VLTIVGRGRRTKLYEAHSDVRESLARIPTIGVQATTCGLYELVEAMVEKGDGSA 1369
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1305 VLTIVGRGRRTKLYEAHSDVRESLARIPTIGVQATTCGLYELVEAMVEKGDGSA 1364
QY 1370 VLELDLCNDRVSRTFFOKCKNKFTTGETTAHGVKGOGISAMSKTFCALFGMPFRAIEKE 1429
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1365 VLELDLCNDRVSRTFFOKCKNKFTTGETTAHGVKGOGISAMSKTFCALFGMPFRAIEKE 1424
QY 1430 IIALLPNIFGYDAYEESVFAAASGASCMVEENDFSEEDSTONNESLGLCECVMEEGC 1489
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1425 IIALLPNIFGYDAYEESVFAAASGASCMVEENDFSEEDSTONNESLGLCECVMEEGC 1484
QY 1490 MPQWLRIRLVASNTILOAKRESLKGEMKSHSGEGTLLMNTVMNAIIAHCEPFEDFR 1549
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1485 MPQWLRIRLVASNTILOAKRESLKGEMKSHSGEGTLLMNTVMNAIIAHCEPFEDFR 1544
QY 1550 VAAFKGDDSVVLCSDYRQSHNAALATAGCGLKLVDRPIGLYAGVVAAGLGLTPDVVR 1609
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1545 VAAFKGDDSVVLCSDYRQSHNAALATAGCGLKLVDRPIGLYAGVVAAGLGLTPDVVR 1604
QY 1610 FAGRLSEKMGPGPERAEQLRLAVCDFLRLTNVAQVCDVVSRYGVSGLVHNLIGML 1669
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1605 FAGRLSEKMGPGPERAEQLRLAVCDFLRLTNVAQVCDVVSRYGVSGLVHNLIGML 1664
QY 1670 QTADGKAHFTETIKPVLDLTNSIIORVE 1698
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1665 QAVADGKAHFTESVKPVLDLTNSIWCVE 1693

RESULT 8
089444 PRELIMINARY: PRT: 1693 AA.
AC 089444;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE UNNAMED PROTEIN PRODUCT.
OS Hepatitis E virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
OX NCBI_TaxID=12461;
RN [1]
RA YIN S.R., Purcell R.H., Emerson S.U.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K52-87;
RX MEDLINE=95176571; PubMed=7871758;
```

```
RA Yin S., Purcell R.H., Emerson S.U.;
RT "A new Chinese isolate of hepatitis E virus: comparison with strains
   recovered from different geographical regions.";
RL Virus Genes 9:23-32(1994).
DR EMBL: L25547; AAA91078.1; -.
DR EMBL: L25595; AAA65488.1; -.
DR InterPro: IPR000606; Viral_helicase1.
DR InterPro: IPR002589; V_methyltransf.
DR pfam: PF01443; Viral_helicase1.
DR pfam: PF01660; V_methyltransf. 1.
DR SMART: SM00506; A1pp: 1.
SQ
SEQUENCE 1693 AA: 185122 MW: 539148302338DSE CRC64:

Query Match      82.7%; Score 7456; DB 12; Length 1693;
Best local Similarity 81.8%; Pred. No. 0;
Matches 1401; Conservative 108; Mismatches 161; Indels 42; Gaps 8;

QY 1 PGTTTAEQAAALAAANSALANAVVREPLSRVOTEIILINMOPROLYFREVLMNPIOR 60
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 10 PGTTTAEQAAALAAANSALANAVVREPLSRVOTEIILINMOPROLYFREVLMNPIOR 69
QY 61 VIHNELEYCARARAGRCLEVGAPRSTINDNPVLRHCFELRPGVDVORWTSAPTRGPAN 120
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 70 VIHNELEYCARARAGRCLEVGAPRSTINDNPVLRHCFELRPGVDVORWTSAPTRGPAN 129
QY 121 CRRSALRGLPAPADRTYCFDFGSRCAFAETGVALYSLHMDLPADVABAMAHRGXTRLYAA 180
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 130 CRRSALRGLPAPADRTYCFDFGSGCNFPAETGIALYSLHMDSPDVABAMRHRGXTRLYAA 189
QY 181 LHLPEVLEPPGYHTTSTYLLHDGDAVYTTESDTAGTNNHDVSLIRAITRTKIYGDH 240
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 190 LHLPEVLEPPGYHTTSTYLLHDGDAVYTTESDTAGTNNHDVSLIRAITRTKIYGDH 249
QY 241 PLVIERVRAIGCHVLLTAPRSPMPVYPVPESTEVYRSIRPGSGSLSPFASCTK 300
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 250 PLVIERVRAIGCHVLLTAPRSPMPVYPVPESTEVYRSIRPGSGSLSPFASCTK 309
QY 301 STEFAPVHIWDRMLMFGATLDDOAFCCSRIMTYLRGISYKVYGVALVANEWNASBDAL 360
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 310 STEFAPVHIWDRMLMFGATLDDOAFCCSRIMTYLRGISYKVYGVALVANEWNASBDAL 369
QY 361 TAYTTAAYLTICHOVRTYRTQAISGMRRLGVEHNAOKITRLYSMLFEKSGRDYIPGQOL 420
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 370 TAYTTAAYLTICHOVRTYRTQAISGMRRLGVEHNAOKITRLYSMLFEKSGRDYIPGQOL 429
QY 421 FYAOCRRMLSGFHLDRVLVFDSEVPCRCRTFLKTVAGKFCGPMRMIGOECTFLPAE 480
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 430 FYAOCRRMLSGFHLDRVLVFDSEVPCRCRTFLKTVAGKFCGPMRMIGOECTFLPAE 489
QY 481 GLVGDHNDNEAYEGSEVDEPAEPAHLDVDSGTYAVHGHQLEALYBALNVPDIIAARSLT 540
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 490 GLVGDHNDNEAYEGSEVDEPAEPAHLDVDSGTYAVHGHQLEALYBALNVPDIIAARSLT 549
QY 541 ATVELVASPDRLECRITVLGNKTFRTTYVDGAHLEANPEOYVLSFDSRSGMGSGHSUT 600
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 550 ATVELVASPDRLECRITVLGNKTFRTTYVDGAHLEANPEOYVLSFDSRSGMGSGHSUT 609
QY 601 YELTPAGLOYRISNGDCATATPPGAGAPSAPEVAFAFCALRYRFRTHSLTJGGLW 660
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 610 YASAGLEVRYVAAGLDHRAVFAFGVSPKSAPEVTAFCALRYRFRTHSLTJGGLW 669
QY 661 LHPGLGIPPPFSPGHIMESANPFCGGLYTRTWS-----TSGFSSDFSPPE 709
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 670 FHPGGLGIPFAPFSPGHIMESANPFCGGLYTRTWSVDAVSSPAQDGLGTFSEPSIPS 729
QY 710 AAPAPMAATPGLPHSTPVSVDIWTLPPEPSEFQVDAAPVPAP--DPAGLPGCVVLTLP 766
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 730 RAA---TLTPAAP-----LPPAP---DPSTPSPAPARGEPA--PG---ATAG 766
QY 767 PPPVHNPSTIPPPSRNRRLTYTPDGAQKYVAGSLFESDCDWLVANASNGHRPGGLCHAF 826
```



```
Db 767 AAPTITQ-----TARHRLFTYPDGSKVFAGSLFESTCTWLVNANVDRHGPGECHAF 821
QY 827 YORFPAFYPTFEREGLAAYTLTPRPITTHAAYPYRVONKRLAEAREFCSRGTA 866
Db 822 YORFPAFSDAASVMDGAAATLTTPRPIITHAAYPDYLEHNFKEAAREFCSRGTA 861
QY 887 AYPPLGSGIYQVPSLSPDAMERNHRPDELYLTPBAAMFEANKPAQVLTITTEDTART 946
Db 882 AYPPLGTGTYQVYIGPSFDMERNHRPDELYLTPBAAMFEANKPAQVLTITTEDTART 941
QY 947 ANLALEIDATVEGRACAGCTISPGIVHQFTAGVSGSKSIOGDDVYVVPFREL 1006
Db 942 ANLALEIDATVDGRACAGCTVPGVYQFTAGVSGSKSITADVDVYVVPFREL 1001
QY 1007 NSMRRGFAFPPTHTAARTYIGRRVIDEAPSLPRLHLLHMO RASSVHLGDPNCPAT 1066
Db 1002 NSMRRGFAFPPTHTAARTYIGRRVIDEAPSLPRLHLLHMO RAKAVHLLGDPNCPAT 1061
QY 1067 DEHAAGLVPAIRBELAPTSMMXVTHRCPADVCELIRGAYPKIQTTSRVLSLFWNBPAT 1126
Db 1062 DEHAAGLVPAIRBDLAPTSMMXVTHRCPADVCELIRGAYPMIQTTSRVLSLFWNBPAT 1121
QY 1127 QKLVYQAANKANPGAIYVHEAGATFETTTIATADAGLIOSRAHATVALTRTEKC 1186
Db 1122 QKLVYQAANKANPGSVTVEHAGATFETTTIATADAGLIOSRAHATVALTRTEKC 1181
QY 1187 VILDAAGLIREVIGISDVYINNFELAGVEGXHRSYIIPGNPNONIGTLOAPPSOISA 1246
Db 1182 VILDAAGLIREVIGISDAIYNFELAGVEGXHRSYIIPGNPNANDYLAAPPSOISA 1241
QY 1247 YHQLAEELGRPAVPAVAVLPCELEQGLLYMPOELVSDVLYVELTJDIVHCRMAPSQ 1306
Db 1242 YHQLAEELGRPAVPAVAVLPCELEQGLLYMPOELVSDVLYVELTJDIVHCRMAPSQ 1301
QY 1307 KRAVSTLVGRGRRTKLYEAHSDVRESLARTPTIGVQATTCCLYELVEAMVEKGD 1366
Db 1302 KRAVSTLVGRGRRTKLYEAHSDVRESLARTPTIGVQATTCCLYELVEAMVEKGD 1361
QY 1367 GSAVLELDLCNDRVSITFEOKCKNFETGETAHGKVGOGISAMSTFCALGPMFRAI 1426
Db 1362 GSAVLELDLCNDRVSITFEOKCKNFETGETAHGKVGOGISAMSTFCALGPMFRAI 1421
QY 1427 EKEIALLPRTNIFYDAYEESVFAAASGAGSCMFENDESEPDSTONNESLGLCEVME 1486
Db 1422 EKEIALLPRTNIFYDAYEESVFAAASGAGSCMFENDESEPDSTONNESLGLCEVME 1481
QY 1487 ECGMPQMLRLYLHVASANTIDAKRESLKGFMKHSGEPTLLMNTVMNAVITTCYDPR 1546
Db 1482 ECGMPQMLRLYLHVASANTIDAKRESLKGFMKHSGEPTLLMNTVMNAVITTCYDPR 1541
QY 1547 DERVAAFKGDSDVILCSDRQSRNAALJAGGLKLVDPRTIGLAGVVAAGLTLDP 1606
Db 1542 DERVAAFKGDSDVILCSDRQSRNAALJAGGLKLVDPRTIGLAGVVAAGLTLDP 1601
QY 1607 VVRFAGRLSEKMMWGPBPBAEOLRLAVCDFLRLGVLNVAOCVDSRVYGVSPGLVHNL 1666
Db 1602 VVRFAGRLSEKMMWGPBPBAEOLRLAVCDFLRLGVLNVAOCVDSRVYGVSPGLVHNL 1661
QY 1667 GMLQTIADGKAHFTETIKRVLDTNSIIO RVE 1698
Db 1662 GMLQTIADGKAHFTETIKRVLDTNSIIO RVE 1693
```

```
RESULT 9
081876 PRELIMINARY: PRT: 1693 AA.
AC 081876:
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
OS COMPLETE GENOME SEQUENCE.
Hepatitis E virus.
```

```
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
OX NCBI_Taxid=12461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UIC8179;
RA Uchida T.;
RL Submitted (MAY-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; D11093; BAA01865.1; -.
DR InterPro: IPR000606; Viral_helicase1.
DR InterPro: IPR002588; V_methyltransf.
DR InterPro: IPR002589; DUF27.
DR Pfam: PF01443; Viral_helicase1.
DR Pfam: PF01660; Vmethyltransf.1.
DR Pfam: PF01661; DUF27.1.
DR SMART; SM00506; A1pp; 1.
SQ SEQUENCE 1693 AA; 184839 MW; 4E422ED55AE0B1 CRC64;

Query Match 82.4%; Score 7424; DB 12; Length 1693;
Best local similarity 81.6%; Pred. No. 0;
Matches 1397; Conservative 108; Mismatches 165; Indels 42; Gaps 8;

QY 1 PGTTTAEQAAIAAANSAANAVVPRFLSRVOTELLINLMOPROLVFRPEVIMNPIQR 60
Db 10 PGTTTAEQAAIAAANSAANAVVPRFLSHOOIELILNLMOPROLVFRPEVIMNPIQR 69
QY 61 VTHNELEOYCRARAGRCLEGAHPRISINDPNVYLRHCFLRPVGRDQVRYSAPTGRPAAN 120
Db 70 VTHNELELYCRARAGRCLEGAHPRISINDPNVYLRHCFLRPVGRDQVRYSAPTGRPAAN 129
QY 121 CRRSALRGLPADRTYCFDGFSCARCAEAEVALYSYSLHMLPDAVMAARHGXTRLYAA 180
Db 130 CRRSALRGLPADRTYCFDGFSCARCAEAEVALYSYSLHMLPDAVMAARHGXTRLYAA 189
QY 181 LHLPEVLLPRTGYHTSTLLIHGDRAYVYBQDTSAGYNHDVSLRLAMIRTKYVGH 240
Db 190 LHLPEVLLPRTGYHTSTLLIHGDRAYVYBQDTSAGYNHDVSLRLAMIRTKYVGH 249
QY 241 PLVIERVRAIGCFHVLTLAAPESSPMYPVPRSTEVYVRSIFGSGSPSLSPSASCSTK 300
Db 250 PLVIERVRAIGCFHVLTLAAPESSPMYPVPRSTEVYVRSIFGSGSPSLSPSASCSTK 309
QY 301 STEFAVPAHIMRLMFGATLDDOAFCCSRLMTYLRGISYKVTYVGLVANEGMNASDAL 360
Db 310 STEFAVPAHIMRLMFGATLDDOAFCCSRLMTYLRGISYKVTYVGLVANEGMNASDAL 369
QY 361 TAXITAAVLTICHORLYRTQALSKGMRRLGVEHAQKFTIRLVSMLFEKSGRDYIPGRQLE 420
Db 370 TAXITAAVLTICHORLYRTQALSKGMRRLGVEHAQKFTIRLVSMLFEKSGRDYIPGRQLE 429
QY 421 FYAOCRRMLASGFHLDPRLVLFDESVPCRTFLKLVAKKFCCEFMRLGOECTCELEPAE 480
Db 430 FYAOCRRMLASGFHLDPRLVLFDESVPCRTFLKLVAKKFCCEFMRLGOECTCELEPAE 489
QY 481 GLVGDHGDNEAYEGSEVDPAEPALDVSCTYAVHSHOLEALYLRALNVQDIAAARSRLT 540
Db 490 GLVGDHGDNEAYEGSEVDPAEPALDVSCTYAVHSHOLEALYLRALNVQDIAAARSRLT 549
QY 541 ATVELVAASPRDLRCRTVLTCTRTVVDGAHLEANGPEOYVLSFASROSQAGASHJT 600
Db 550 ATVELVAASPRDLRCRTVLTCTRTVVDGAHLEANGPEOYVLSFASROSQAGASHJT 609
QY 601 YELTPAGLOVRISSNGLDCTATFPFGAGPAPAGVAAACSAALYRYNRTQHSRLTGLAM 660
Db 610 YELTPAGLOVRISSNGLDCTATFPFGAGPAPAGVAAACSAALYRYNRTQHSRLTGLAM 669
QY 661 LHPGGLLTGTPPSPGHITWESANPCGEGTLYTRTWS-----TSGSSDSPPPE 709
Db 670 LHPGGLLTGTPPSPGHITWESANPCGEGTLYTRTWS-----TSGSSDSPPPE 729
QY 710 AAAPMAAATPGVPHSPRSDIWLVPPESEEOVAAAPRPP---DPAGLGPVVLTPP 766
Db 730 RAA---ILTPAP-----LPPAP---DPSPTPSAPARGEPA--PG---ATAG 766
```


QY	767	PPPPVHKPSITPPSRNRKLLIYYTPROGAKYVAGSLFESQDMLVYANSNGHPRGGGLCHAF	826
Db	767	APAITIQ-----TARRHRLFTFYPPGSKVFASLESTCYTMVYANSNDHPRGGGLCHAF	821
QY	827	YQAFPEAFYPTFEEIMREGIAAATLTLPRLPIIAHVADYRVEQPKRLFAARETCSRCTA	886
Db	822	YQYFASPFDAASVYMDGAAATLTLPRLPIIAHVVDYRLFNHPKRLDEAAYRETCSRCTA	881
QY	887	AYVLLGSGIYQVPSLSEFDAMERNHRPGDELYLTERPAANWEFANKPAQVLTTEDEART	946
Db	882	AYPLGTGTYQVYIGFSPFAMERNHRPGDELYLTERLAAMEFANKPCTPLTTEDEVART	941
QY	947	ANLALFEDAETEGRACACCTISPGIVHQAQFAGVPGSKSSTIOGGVDVYVYPTRELR	1006
Db	942	ANLALFEDATDIDGRACACCRVTPGVQYQFPAQVPGSKSSTIQADVDDVYVYPTRELR	1001
QY	1007	NSMRRGGAFAFTPHTAARTITGRVVYIDAPSLPHLLLLHHQRASSVHLLGDPNQIPAI	1066
Db	1002	NAMRRRGGAFAFTPHTAARTYTGGRVYIDAPSLPHLLLLHHQRAATVHLLGDPNQIPAI	1061
QY	1067	DEFHAGLVPAIRPELAPTSQWXYVTHRCPADVCELLRGAVPKIQTTSRVLRSLFWNEPAIG	1126
Db	1062	DFEHAGLVPAIRBDLAPTSQWXYVTHRCPADVCELLRGAVPMIQTTSRVLRSLFWNEPAVG	1121
QY	1127	QKLYVXQAKAANPAGSITVTHEOGAFAFTEETITIAADAGLGIQSSRAHAIVALTPHTEKC	1186
Db	1122	QKLYFTQAKAANPAGSVTTHEOGATYVTEETITIAADAGLGIQSSRAHAIVALTPHTEKC	1181
QY	1187	VILDAPGLRENGISDVIVNNFELGAGEVGHXRPSTYIPKGNPDQMLGTLOAFPPSCQISA	1246
Db	1182	VIIDAPGLVREVAISDAIVNNFELGAGEIGHQRPSTYIPKGNPDANDVDTLAFPPSCQISA	1241
QY	1247	YHOLAEELGHRPAPVAAYVPCPELEQGLLYMPQELTVSDSVLYVELTDIVHCRMAAPSQ	1306
Db	1242	FHOLAEELGHRPAPVAAYVPCPELEQGLLYVPOELTTCDSVYVTEFLDVIHCRMAAPSQ	1301
QY	1307	RKAVLSTLVGRYGRPKLTLEAAHSDVRESLARIPIPIGVQATQCEIYELVEAMVEKQD	1366
Db	1302	RKAVLSTLVGRYPAGTKLTINASHSDVRSDLARIPRIGPVQYTTCEIYELVEAMVEKQD	1361
QY	1367	GSAVLELIDLCNRDVSRIITEFFOKXCNKFTTGETIAGHKVQGISAMSKTFICALFGWPRAI	1426
Db	1362	GSVAVLELIDLCNRDVSRIITEFFOKXCNKFTTGETIAGHKVQGISAMSKTFICALFGWPRAI	1421
QY	1427	EKETLALIPNITVGDATVEESVPAANAVSAGSACMPFENDFSEFSDTONNFSIGLECYME	1486
Db	1422	EKATLALLPGGVYFGADFDYVESAANAAKSMYFENDFSEFSDTONNFSIGLECAJME	1481
QY	1487	ECGAMPWLILYLVLRVSAMTLOPKRSTLGFPMKKSISGEGTLLMTVMNMAITIAHCYFR	1546
Db	1482	ECGMPWLILYLVLRISAMTLOPKRSTLGFPMKKSISGEGTLLMTVMNMAITIAHCYFR	1541
QY	1547	DFRYAAAFKGGDSVVLCSDYRQSHNAALJAGGLKLIKVDYRIGLYAGVYVAPGIGTLPLD	1606
Db	1542	DLQVAAAFKGGDSVYLCSERYQSPGAAYVLLJAGGLKLIKVDYRIGLYAGVYVAPGIGTLPLD	1601
QY	1607	VYVFAAGLSTKNNGPBPBPAPAEQRLAVCPFLGLTINAVQCYDVYSRYVGSPLGVNHLI	1666
Db	1602	VYVFAAGLSTKNNGPBPBPAPAEQRLAVSOFELKRLTNVAMQCYDVYSRVYVGSPLGVNHLI	1661
QY	1667	GMLQTLADGKAFHETETIKPLDLTNSIIDRVE	1698
Db	1662	GMLQVADGKAFHETESYKPYVLDLITNSIIDRVE	1693
RESULT	10		
ID	081862	PRELIMINARY:	PRT: 1693 AA.
AC	081862:		
DT	01-NOV-1996 (Tremblrel, 01, Created)		
DT	01-NOV-1996 (Tremblrel, 01, Last sequence update)		
DT	01-JUN-2001 (Tremblrel, 17, Last annotation update)		

DE ORF 1.
OS Hepatitis E virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
OX NCBI_TaxId=12461;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=92024067; PubMed=1926770;
RA Tam A.W., Smith M.M., Guerra M.E., Huang C.C., Bradley D.W., Fry K.E.,
RA Reyes G.R.;
RT "Hepatitis E virus (HEV): molecular cloning and sequencing of the
RT full-length viral genome.";
RL Virology 185:120-131(1991).
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE=92261377; PubMed=1584074;
RA Uchida T., Suzuki K., Hayashi N., Iida F., Hara T., Oo S.S.,
RA Wang C.K., Shikata T., Ichikawa M., Rikihisa T., Mizuno K., Win K.M.;
RT "Hepatitis E virus: cDNA cloning and expression.";
RL Microbiol. Immunol. 36:67-79(1992).
RN (3)
RP SEQUENCE FROM N.A.
RX MEDLINE=9235008; PubMed=1630924;
RA Aye T.T., Uchida T., Ma X.Z., Iida F., Shikata T., Zhuang H.,
RA Win K.M.;
RT "Complete nucleotide sequence of a hepatitis E virus isolated from the
RT Xinjiang epidemic (1986-1988) of China.";
RL Nucleic Acids Res. 20:3512-3512(1992).
RN (4)
RP SEQUENCE FROM N.A.
RX MEDLINE=92115700; PubMed=1711327;
RA Tsarev S.A., Emerson S.U., Reyes G.R., Tsareva T.S., Legters L.J.,
RA Malik I.A., Iqbal M., Purcell R.H.;
RT "Characterization of a prototype strain of hepatitis E virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:559-563(1992).
RN (5)
RP SEQUENCE FROM N.A.
RX MEDLINE=92271462; PubMed=1589964;
RA Fry K.E., Tam A.W., Smith M.M., Kim J.P., Luk K.C., Young L.M.,
RA Platak M., Feldman R.A., Yun K.Y., Purdy M.A., Et Al.;
RT "Hepatitis E virus (HEV): strain variation in the nonstructural gene
RT region encoding consensus motifs for an RNA-dependent RNA polymerase
RT and an ATP/GMP binding site.";
RL Virus Genes 6:173-185(1992).
RN (6)
RP SEQUENCE FROM N.A.
RX BI S.L., Purdy M.A., McCaustland K.A., Margolis H.S., Bradley D.W.;
RA "The sequence of hepatitis E virus isolated directly from a single
RT source during an outbreak in China.";
RL Virus Res. 28:233-247(1993).
RN EMBL; L08816; AAA03189.1;
DR InterPro: IPR000606; Viral_helicase1.
DR InterPro: IPR002588; V_methyltransf.
DR InterPro: IPR002589; DUF27.
DR Pfam: PF01443; Viral_helicase1. 1.
DR Pfam: PF01660; Vmethyltransf. 1.
DR Pfam: PF01661; DUF27. 1.
DR SMART: SM00506; Altp. 1.
SQ SEQUENCE 1693 AA; 185052 MW; D3363602435B8FF5 CRC64;

Query Match	82.28;	Score 7412;	DB 12;	Length 1693;
Best Local Similarity	81.78;	Pred. No. 0;		
Matches 1395;	Conservative 106;	Mismatches 174;	Indels 32;	Gaps
Oy	1	PGITTAEQALAAANSALANAAVVPRLSVQTEILINLMQPOLVFREVEVLNHP	IQR	60
Db	10	PGITTAIQGALAAANSALANAAVVPRLSVQTEILINLMQPOLVFREVEVLNHP	IQR	69
Oy	61	VIHNELEQYCARARGCLCEVGAHPRSINDNPVNLHRCFLRPVGQDVQRYTSAPTRGPAAN		120
Db	70	VIHNELEQYCARARGCLCEIHAHPRSINDNPVNVHRCFLRPAGQDVQRYMTAPTRGPAAN		129

QY	121	CRNSLRCLPRADPTTCYCFDGFSCRAFAEYGVALLYSLHMDLPADVAEAMARHGTRLYAA	180
Db	130	CRSRSLRCLPADPTTYTCFDFSGCNFAFEYGVALLYSLHMDSPDVAEAMFRRGMRLXYAA	189
QY	181	LHLPEVLLPGCYHTHTSYLLIHGDGAVVYTGCDTSAGYNHDVSLTRAMITRTKLYGDH	240
Db	190	LHLPEVLLPGCYRTASLYLLIHGGRVVYTTGCDTSAGYNHDVSLNRSMITRTKYTGDH	249
QY	241	PLVIERVRAICHEFVLLLTAAPEBSPMPVYVPRSTEVYVRSJFGBGSPSLFPGASCSTK	300
Db	250	PLVIERVRAICHEFVLLLTAAPEBSPMPVYVPRSTEVYVRSJFGBGSPSLFPGASCSTK	309
QY	301	STFNAHVPYHIDRLMFGATLDQAFCCSCLMYTLLNGISIKYTVGALVANEQWANSBDAL	360
Db	310	STFNAHVPYHIDRLMFGATLDQAFCCSCLMYTLLNGISIKYTVGTLVANEQWANSBDAL	369
QY	361	TAXITAAVLTTCCHORYLRTQASISGMRLGVEHAQFRTIRLSWLPKSGRGYTPGRQLE	420
Db	370	TAVITAAVLTTCCHORYLRTQASISGMRLREHAQFRTIRLSWLPKSGRGYTPGRQLE	429
QY	421	FYAQCRMLWSAGFHLIDRPNVLFDESVYRCRTEFLKLVAGKFCCEFMRLGDEQTCFLPEA	480
Db	430	FYAQCRMLWSAGFHLIDRPNVLFDESBACHCRTLIRAVKSFCECFMRLGDEQTCFLQPAE	489
QY	481	GLVGDHGDNDATATGSEYDPAEPALHDVSTYVNHGQLEALYRALNPQDIAARSLT	540
Db	490	GAVGGGHDNDATATGSDVDPAESAISDISYVVPSTALQPLVQADLPAEIVAAARSLT	549
QY	541	ATVELVAPSDLECRVYLGKNTKPTTYVVDANLEANGPBOVYVSPFASQSGASHSLT	600
Db	550	ATVYKSYQDNGIIDEETLLGNKTFRTSTSYVDQAVLETNGPERHNLSPFASQSTMAAPSLT	609
QY	601	YELTPAGLQVRISSNGIDCTATPPEGAPSAABEGEVAACFSCALYRYNRPTORHSLTGLW	660
Db	610	YASASAGLEVRYVAAGLDHRVAFRAGVSPRSABGEVATFACSALYRPNRBAQSLSLTGNW	669
QY	661	LHPGGLIGTFPPSPRGHIMESANPFCGEGLLYTRTMS-TSGSSDSP-----PRAAPA	714
Db	670	FHPBGLLGFPAFSPRGHIMESANPFCGEGLLYTRTMSVDVAVSSPAPDPLGFIISPSIPS	729
QY	715	MAAPRGLPHSPRPVSDIMVLPPESEFQVOVAAV-PPADPQAGLP-GRVLPFRPPRPVH	772
Db	730	RAAT-----LTP-----AALPPRADPSPRTSPSAPARGBAPGATAR	766
QY	773	KPSIP-PPSRNRRLLYTPPDGAKYVAGSLFESDCDMLVNASNPGRNRPGGGLCHAEYOREP	831
Db	767	APALTHQAARRRLLFYTPDGSKVFGSLFESPCTMLVNASNVDRHPGGGLCHAEYQYRP	826
QY	832	EAFPTPEITMEGLAAVTLPRPLIHAVADYVEQDNPKRELANRETCSRGTATAPLL	891
Db	827	ASFPASASVMDGAAATVTLPRPLIHAVADYLEHNPKMLEAAVRETCSRGTATAPLL	886
QY	882	GSGIYOVNVSLSFPAEMRNRPBGDELYLTPRANWFEAKKPAQVYLTTEEDTAPRANLAL	951
Db	887	GTGIIYQVPIGCSFPAEMRNRPBGDELYLTPRLAKWFEAKKNPCPFLTTEEDTAPRANLAI	946
QY	952	EIDATEVGRACACSTSPGIVHYOFTAGYPGSGKRSRISQCGDVVVVVPRTELNSMR	1011
Db	947	ELDSATDVGRCACACRYTPGVYQYGFAGYPGSGKRSRISTQADVVVVVVPRTELNMR	1006
QY	1012	RGEPAFTPHTAARVLTIGRRVVIDEAPSLPHLLLLHMORASSVHLGDPNQTALDFEHA	1071
Db	1007	RGEPAFTPHTAARVLTIGRRVVIDEAPSLPHLLLLHMORAAVHLLGDPNQTRALDFEHA	1066
QY	1072	GLVPAIRBELAPRSMWXYTHRCRADVCCELLRGAYPKIOTTSRYVLSLFPNBEAIOQLVX	1131
Db	1067	GLVPAIRBDLAPRSMWXYTHRCRADVCCELLRGAYPMIOTTSRYVLSLFWGEPVQOKLVF	1126
QY	1132	TOAKKANPGATVYHNEOGATFETETIATADARGLIOSSRAHAIVALLRHEKEVILDA	1191
Db	1127	TOAKKANPGSVYHNEOGATYETETIATADARGLIOSSRAHAIVALLRHEKEVILDA	1186
QY	1192	PGLLRVGISDVIVNNEFLAGGEVGHKRPVIRGNBDQNLGTLOAFPPSCISAYHDLA	1251

Query	Match	81.4%: Score 7333; DB 12; Length 1693;	Best Local Similarity 79.8%; Pred. No. 0;	Matches 1375; Conservative 120; Mismatches 163; Indels 66; Gaps 6
QY	1	PEITTAIEQAALAAANSAALANAVVAPPELSRYQTETLLIMQPROLVFPEVLMNHPOR	60	
DB	10	PEITTAIEQAALAAANSAALANAVVAPPELSRYQTETLLIMQPROLVFPEVLMNHPOR	69	
QY	61	VIHNELEOYCRARAGCLEVGAHAPSINDPNVLRHCFILPVGGRDVOYRYSAPTRGPAA	120	
DB	1187	PGLLREVGISDIALVNNFETLAGGEIGHQRSSVILPRGPDANVDTLAAPPSCQISAHQIA	1246	
QY	1252	EELGHPAPVAANAALPCCPELEOGLTAMPDELTVSDSLVFELTDIYHGMMAAPSQRAVL	1311	
DB	1247	EELGHPAPVAANAALPCCPELEOGLTAMPDELTVSDSLVFELTDIYHGMMAAPSQRAVL	1306	
QY	1312	STLVGRYGRRTKLYEAHSDVRESLARFTPTIGPVQATTCELYELVEAVENKQDGSAYL	1371	
DB	1307	STLVGRYGRRTKLYEAHSDVRESLARFTPTIGPVQATTCELYELVEAVENKQDGSAYL	1366	
QY	1372	EILDICRQVSRITTEPOKXONKFTTGTTIAGKVGQGISAMSKPFCALPMPRAIEKELL	1431	
DB	1367	EILDICRQVSRITTEPOKXONKFTTGTTIAGKVGQGISAMSKPFCALPMPRAIEKELL	1426	
QY	1432	ALLPNIPIFYDAVEESVFAAASGASCVWFENDESEPTSTONNFTSLGECVMEECGAP	1491	
DB	1427	ALLPQVFFVGDAADDTVFSAVAANAASVNFENDESEPTSTONNFTSLGECALMVECGAP	1486	
QY	1492	QMLIRLYHLVRSAMILQAPRESLKGFWKHSGEPTLLMNTVNMMAIIAHCEEFDRFYA	1551	
DB	1487	QMLIRLYHLVRSAMILQAPRESLKGFWKHSGEPTLLMNTVNMMAIIAHCEEFDRFYA	1546	
QY	1552	ARKGDSVYLCSDYRQSRMAALLIAGCGGLKLVYDYPIDGLYAGVVAAPGLGTLPPDVYRA	1611	
DB	1547	ARKGDSVYLCSDYRQSRMAALLIAGCGGLKLVYDYPIDGLYAGVVAAPGLGTLPPDVYRA	1606	
QY	1612	GRLEKNMGPGGERAQLALVACDFIRGLTNAQVVDVAVSRVYGGSPGLVHNLIGMLQT	1671	
DB	1607	GRLEKNMGPGGERAQLALVACDFIRGLTNAQVVDVAVSRVYGGSPGLVHNLIGMLQT	1666	
QY	1672	IADKAHFTETIKPVLDTNLSIIQVE	1698	
DB	1667	IADKAHFTETIKPVLDTNLSIIQVE	1693	
RESULT	11			
Q9WC28		PRELIMINARY; PRT; 1693 AA.		
AC	Q9WC28;			
DT	01-NOV-1999 (TREMBlrel. 12, Created)			
RT	01-NOV-1999 (TREMBlrel. 12, Last sequence update)			
RL	Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AF076239; AAC27934.2; -			
DR	InterPro: IPR000606; Viral_helicase1.			
DR	InterPro: IPR002588; V_methyltransf.			
DR	InterPro: IPR002589; DUF27.			
DR	Pfam: PF01443; Viral_helicase1.1.			
DR	Pfam: PF01660; Urmethyltransf.1.			
DR	Pfam: PF01661; DUF27.1.			
DR	SMART: SM00506; Alpp.1.			
SO	SEQUENCE 1693 AA; 185578 MW; FB37B87BF6BB501 CRC64;			

```
Dh 70 VINHELELCARSGCILEGAHPRSLINDPNVNHRCFLRPVGRDVGWYTAPTRGPAAN 129
Qy 121 CRSSALRGLEPPADRTYCFDGFSCAPAEETGVALYSLIHDLMPADVAEAMAHGXTRLYYA 180
Dh 130 CRSSALRGLEPPADRTYCFDGFSCAPAEETGVALYSLIHDLMPADVAEAMAHGXTRLYYA 189
Qy 181 LHLPEVLLPQGYHTTSYLLIHDGRVAVTYEGDTSAGYNHDSYSLRAIRTKTYGDH 240
Dh 190 LHLPEVLLPQGYHTTSYLLIHDGRVAVTYEGDTSAGYNHDSYSLRAIRTKTYGDH 249
Qy 241 PLVIEVRAGCHFVLLTAAPESPMPYPVPRSTEVYRSIFGPGSGSLTPSACSTK 300
Dh 250 PLVIEVRAGCHFVLLTAAPESPMPYPVPRSTEVYRSIFGPGSGSLTPSACSTK 309
Qy 301 STFHAHVHIMDRMLFGATLDDQAFCCSRLMTYLRGISYKVTGALVANEGNASDAL 360
Dh 310 STFHAHVHIMDRMLFGATLDDQAFCCSRLMTYLRGISYKVTGALVANEGNASDAL 369
Qy 361 TAITTAAYLTICHORYLRTQALISGMHRLGVEHAOKFITLTYSLPEKSGRDYTPGRQLQ 420
Dh 370 TAITTAAYLTICHORYLRTQALISGMHRLGVEHAOKFITLTYSLPEKSGRDYTPGRQLQ 429
Qy 421 FYAQCRRLMSAGHLDPRVLPFDESVPQRCRTFLKVAAGFCFEMRLGQECFLEPAE 480
Dh 430 FYAQCRRLMSAGHLDPRVLPFDESVPQRCRTFLKVAAGFCFEMRLGQECFLEPAE 489
Qy 481 GLVGDHNDNEAYEGSEVDAEPAHLDVSGTYAVHGQLEALYRALNPODILARASRLT 540
Dh 490 GLVGDHNDNEAYEGSEVDAEPAHLDVSGTYAVHGQLEALYRALNPODILARASRLT 549
Qy 541 ATVEIYASPRLECRVYLGKTKRTTYVVDGAHLEANGPEQYVLSFDMSRQSMAGSISLT 600
Dh 550 ATVEIYASPRLECRVYLGKTKRTTYVVDGAHLEANGPEQYVLSFDMSRQSMAGSISLT 609
Qy 601 YELTPAGLOYRISNGLDCAATFPFGGAPSAAPGEVAFSALYRYRNFQORSLTGLW 660
Dh 610 YELTPAGLOYRISNGLDCAATFPFGGAPSAAPGEVAFSALYRYRNFQORSLTGLW 669
Qy 661 LMBEGGLIFPPSPGHIMESANPFCGEGTLTYRTWS-----TSGFSSDFSPPE 709
Dh 670 FHEBGLIGLEFAPSPGHIMESANPFCGEGTLTYRTWS-----TSGFSSDFSPPE 726
Qy 710 AAAPMAATPGLPHSTIPVSDIVLPPSEFQVDAAPVPAPDPAGLPGPVVLTLPPPP 769
Dh 727 - -TPSRAPPTL-----AAPLPP-----LAPDPS 749
Qy 770 PVHKPSIAPP-----SRRRLTYTPDGAAVYVAGSLFESDCDMLVNASNP 814
Dh 750 PSSAPALDEPASATSGVPAITHTQTAHRRLLFTYPDGSKVFAGSLFESTCTMLVNASNP 809
Qy 815 GHRPGLGCHAFYQRPPEAFYPTFEIMREGIAAYTLTPRIIHAVADYVNEQNPRLFA 874
Dh 810 DHCPPGLGCHAFYQRPPEAFYPTFEIMREGIAAYTLTPRIIHAVADYVNEQNPRLFA 869
Qy 875 AYETSGSRKTAAYPLIGSSITQYVPSLSFDANERNRPDEDELYLTPAANPEANKPAQ 934
Dh 870 AYETSGSRKTAAYPLIGSSITQYVPSLSFDANERNRPDEDELYLTPAANPEANKPAQ 929
Qy 935 PVLTIEDTARTANLAEIDAATEVGRACAGCTISPGIYVQFAGVPGSGSKSITQGD 994
Dh 930 PVLTIEDTARTANLAEIDAATEVGRACAGCTISPGIYVQFAGVPGSGSKSITQGD 989
Qy 995 VDVVVVPTRELNRSMRRGFAFTPHTAAVTTIGRRVVIDEASLPRPHLLLHMQRASV 1054
Dh 990 VDVVVVPTRELNRSMRRGFAFTPHTAAVTTIGRRVVIDEASLPRPHLLLHMQRASV 1049
Qy 1055 HLLGDPQIATIDFENAGLVPAIRPELAPTSWAXVTHRCADVCCELLRGAYPKIQTTSRV 1114
Dh 1050 HLLGDPQIATIDFENAGLVPAIRPELAPTSWAXVTHRCADVCCELLRGAYPKIQTTSRV 1109
Qy 1115 LRSLEWNEPAGIOLVYTOAKAANPGATVHEOGATFETETITATADRGILQSSRAH 1174
Dh 1110 LRSLEWNEPAGIOLVYTOAKAANPGATVHEOGATFETETITATADRGILQSSRAH 1169
```

```
Qy 1175 AYAALRHTKECYITLDAPGLLREVGISDYVYNNFFLAGGEVGHRSVPVIRGNPDNLCT 1234
Dh 1170 AYAALRHTKECYITLDAPGLLREVGISDYVYNNFFLAGGEVGHRSVPVIRGNPDNLCT 1229
Qy 1235 LOAFPSCQISAYHOLAEEIGHRPAPAAVLPCCPELEOGLVMPQELTVSDSLVELT 1294
Dh 1230 LOAFPSCQISAYHOLAEEIGHRPAPAAVLPCCPELEOGLVMPQELTVSDSLVELT 1289
Qy 1295 DIVCHMAAPSORKAVSLTVGRYGRRTKLYEAHSDVRESLARFIPTTIGVOATTCELY 1354
Dh 1290 DIVCHMAAPSORKAVSLTVGRYGRRTKLYEAHSDVRESLARFIPTTIGVOATTCELY 1349
Qy 1355 ELVEAMVEKQDGSAYLELDLCNRDYSRTFPOKXCKFTTGETIANGKVGQGISAMSKT 1414
Dh 1350 ELVEAMVEKQDGSAYLELDLCNRDYSRTFPOKXCKFTTGETIANGKVGQGISAMSKT 1409
Qy 1415 FCALFGFWFAIKEITALLPNIIFYGDAAEESVFAAAGSAGCWFEEDFSEPDSTON 1474
Dh 1410 FCALFGFWFAIKEITALLPNIIFYGDAAEESVFAAAGSAGCWFEEDFSEPDSTON 1469
Qy 1475 NFSIGLECVVMEBEGMPQWLIRLYHLVRSAMIIQAPKESILKGFMKHSGEPGLLMTVW 1534
Dh 1470 NFSIGLECVVMEBEGMPQWLIRLYHLVRSAMIIQAPKESILKGFMKHSGEPGLLMTVW 1529
Qy 1535 NMAIIACHYEFDPFVAFAFGDSDVVLCSQYROSRRNAALLIAGCGKLKYDPRIGLYAG 1594
Dh 1530 NMAIIACHYEFDPFVAFAFGDSDVVLCSQYROSRRNAALLIAGCGKLKYDPRIGLYAG 1589
Qy 1595 VVAAPGIGTLPDVYRFAPGRSEKNMGPGPERAQLRLAAYCDPLRGLTNVNOVCVDVYSRV 1654
Dh 1590 VVAAPGIGTLPDVYRFAPGRSEKNMGPGPERAQLRLAAYCDPLRGLTNVNOVCVDVYSRV 1649
Qy 1655 YGVSPGLVHNLIGMLQTIADGKAHFTETIRKPVLDLTSIIQRYE 1698
Dh 1650 YGVSPGLVHNLIGMLQTIADGKAHFTETIRKPVLDLTSIIQRYE 1693
```

```
RESULT 12
ID 039221 PRELIMINARY; PRT; 1693 AA.
AC 039221:
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE NONSTRUCTURAL POLYPEPTIDE.
OS Hepatitis E virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
ON NCBI_taxid=12461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HYDERABAD, INDIA;
RA Ansari I.H., Manda S.K., Durgapal H., Jameel S., Panda S.K.;
RT "Eukaryotic expression of nonstructural protein (ORF1) of HEV: absence
RT of any processing.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
EMBL: AF028091; AAB82002.2; -
DR InterPro: IPR000606; Viral_helicase1.
DR InterPro: IPR002588; Viral_helicase1.
DR InterPro: IPR002589; DUF27.
DR Pfam: PF01443; Viral_helicase1; 1.
DR Pfam: PF01660; Vmethyltransf; 1.
DR SMART: SM00506; Alpp; 1.
KW Polyprotein.
SQ
SEQUENCE 1693 AA; 185562 MW; 7AFED003532E12BC3 CRC64;
```

```
Query Match 81.3%; Score 7329; DB 12; Length 1693;
Best Local Similarity 79.7%; Pred. No. 0;
Matches 1374; Conservative 122; Mismatches 163; Indels 66; Gaps 6;
Qy 1 PGITTAIEQAAALAAANSALANAVVVRFFLSRVQTEILINMQPQVLVRFREVLMMHPIQR 60
```

```

10 PGITVEGAALATANSALANAVVRPFLSHQIELLINLMOPROLVREVEWMNP10R 69
61 VINELEOYCRARAGCLEVGAHPRISINDPNVLRHCFLEPRVDORVMYSAPTRGPAAN 120
70 VINELELCRAASRGCLLEGAHPRISINDPNVVRHCFLEPRVDORVMYSAPTRGPAAN 129
121 CRRSALRGLEPADRYVCDFGFSRCAPAEATGVALYSLHDLMPADVAEAMARHGXTLYAA 180
130 CRRSALRGLEPADRYVCDFGFSRCAPAEATGVALYSLHDLMPADVAEAMARHGXTLYAA 189
181 LHLPEVLLPBGTYHTTSTVLLJHGDRAVYVTEGDTSAGINHDSVNLRSWIRTKYTGCH 240
190 LHLPEVLLPBGTYHTTSTVLLJHGDRAVYVTEGDTSAGINHDSVNLRSWIRTKYTGCH 249
241 PLVIEVRALICGHFVLLTLAAPERSPMPYVPRSTEVYRSIFGSGSPLEPASCSTK 300
250 PLVIEVRALICGHFVLLTLAAPERSPMPYVPRSTEVYRSIFGSGSPLEPASCSTK 309
301 STFHAVPVHIDRLMLFGATLDDOAFCCSRMLTYLRGISYKVTGVALVANEGNASDAL 360
310 STFHAVPVHIDRLMLFGATLDDOAFCCSRMLTYLRGISYKVTGVALVANEGNASDAL 369
361 TXXITFAVLTICORVLRTOALSKGMRRLGVEHAKOETRLYSWLEKSGROVTPROLO 420
370 TAVITFAVLTICORVLRTOALSKGMRRLGVEHAKOETRLYSWLEKSGROVTPROLO 429
421 FVAOGRMLTLAGHLDPRVLVEDESVPCHRTFLKVKAGKFCFMMRLGOECTCFLEPAE 480
430 FVAOGRMLTLAGHLDPRVLVEDESVPCHRTFLKVKAGKFCFMMRLGOECTCFLEPAE 489
481 GLVGDHNEAVEESVDPAPBAHLVDVGTAVHGHOLEALYRALNVPDIAARSLT 540
490 GYVGDGHNEAVEESVDPAPBAHLVDVGTAVHGHOLEALYRALNVPDIAARSLT 549
541 AATVEIASPDRLCETRVLCNKTRFTVVDGAHLFANGPPOVYLSFSPASQSGASHLT 600
550 AATVEIASPDRLCETRVLCNKTRFTVVDGAHLFANGPPOVYLSFSPASQSGASHLT 609
601 YELTPAGLOVRISNGLDCTATFPFGAPSAAPGEVAFCASALYRNRTQHRSLTGLAM 660
610 YASAGLEVRVYGAOLDHRAIFAPGVSPRNSNGEYATACSAALYRNREAGHRSLTGLAM 669
661 LHPBGLLGFPPSPGHITESANPFCGECTLYTRWS-----TSFSSDPSPE 709
670 FHPBGLLGFPPSPGHITESANPFCGECTLYTRWS-----TSFSSDPSPE 726
710 AARAPAAATPGLPHSTPRVSDIMVLRPPESEFOVDAAPVPAPDAGLGPVLTTPPPP 769
727 --TPSRAATPVL-----AAPLP-----LADPSP 749
770 PVHKPSIPPP-----SRNRLLTYTPDGAKEYAGSLFESDCMLVNASNP 814
750 PSSAPALDPRASAATSGVPAITHOTARHRRLLFTYDGSKAVAGSLFESTCMLVNASNP 809
815 GHPGGGLCHAFYQREPEAFYPTERTMEGLAAYTLTPRPIIHAAVADYRVBONPKRLDA 874
810 DHPGGGLCHAFYQREPEAFYPTERTMEGLAAYTLTPRPIIHAAVADYRVBONPKRLDA 869
875 AATRECSRGSTAYPRLGSGIYQVPYSLFSDAMERNHRGDELYLTERPANNFEAKPKAO 934
870 AATRECSRGSTAYPRLGSGIYQVPYSLFSDAMERNHRGDELYLTERPANNFEAKPKAO 929
935 PLUTTEDPARTANLALTEIDATEVGRACAGCTISPGIYHYOFTAGVPSSGSRSTIOGD 994
930 PLUTTEDPARTANLALTEIDATEVGRACAGCTISPGIYHYOFTAGVPSSGSRSTIOGD 989
995 VDVVVVPTRELNSMRNRGFAAFTPHTAARVLTIGRRVVIDEAPSLPHLLLHMGRASSY 1054
990 VDVVVVPTRELNSMRNRGFAAFTPHTAARVLTIGRRVVIDEAPSLPHLLLHMGRASSY 1049
1055 HLGDPNOIPALIDFEPHAGLVPAIRBELAPTSMWXYTHRCADVCELIRGAVYKIOGTSRY 1114

```

```

1050 HLGDPNOIPALIDFEPHAGLVPAIRBELAPTSMWXYTHRCADVCELIRGAVYKIOGTSRY 1109
1115 LSLFNEPEALIGOKLVTXTOAKANPAGTIVHEAGATFETETIATADARGLIOSSRAH 1174
1110 LSLFNEPEALIGOKLVTXTOAKANPAGTIVHEAGATFETETIATADARGLIOSSRAH 1169
1175 AVALTRHTEKCVIILAPGLREVGISDYIVNFFLAGEVGHXHPVSIKPNDOILGT 1234
1170 AVALTRHTEKCVIILAPGLREVGISDYIVNFFLAGEVGHXHPVSIKPNDOILGT 1229
1235 LOAFPPSCIOISAVHOLAEELGHRPAPVAALVPPCPLEBGLLYMPELIVSSVYFELT 1294
1230 LOAFPPSCIOISAVHOLAEELGHRPAPVAALVPPCPLEBGLLYMPELIVSSVYFELT 1289
1295 DIVHCRMAAPSOGRKAVLSTLVGRYGRRTLYEAHSDVESLAREIPTIGPVQATTCELY 1354
1290 DIVHCRMAAPSOGRKAVLSTLVGRYGRRTLYEAHSDVESLAREIPTIGPVQATTCELY 1349
1355 ELVEAMVERKODGSAAVLELDLCNRDVSRTTFFOKXCNKFTTGTTAAGVAGGISAMSKT 1414
1350 ELVEAMVERKODGSAAVLELDLCNRDVSRTTFFOKXCNKFTTGTTAAGVAGGISAMSKT 1409
1415 FCALGEPMPRAIEKELLALPNIPIFYGDAYEESVFAAAYSGASGCVFENDESSEPSTON 1474
1410 FCALGEPMPRAIEKELLALPNIPIFYGDAYEESVFAAAYSGASGCVFENDESSEPSTON 1469
1475 NFSLECEVMECEGMPQWILRLYLVRSAWILQAPKESLKGFWKKHSGEPGLTMMTW 1534
1470 NFSLECEVMECEGMPQWILRLYLVRSAWILQAPKESLKGFWKKHSGEPGLTMMTW 1529
1535 NNAITIAHCEYEDFRVAERKGDSDVYLCSDYROSNAALLAGCGKLKVDYRPIGLYAG 1594
1530 NNAITIAHCEYEDFRVAERKGDSDVYLCSDYROSNAALLAGCGKLKVDYRPIGLYAG 1589
1595 VVVAPELGTLPVVRPAGRLSEKNMGSPERPEALQALACDPLRGLTNAOVCVDVVSRY 1654
1590 VVVAPELGTLPVVRPAGRLSEKNMGSPERPEALQALACDPLRGLTNAOVCVDVVSRY 1649
1655 YGVSPGLVNLIGMLQTIADGRAHFTETIKPVLDTNSIIORVE 1698
1650 YGVSPGLVNLIGMLQTIADGRAHFTETIKPVLDTNSIIORVE 1693

RESULT 13
081344
ID 081344 PRELIMINARY; PRT: 1693 AA.
AC 081344:
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE HEPATITIS E VIRUS COMPLETE GENOME.
OS Hepatitis E virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
OX NCBI_Taxid=12461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HEBEI;
RX MEDLINE=93348763; PubMed=8346669;
RA Bi S.L., Purdy M.A., McCausland K.A., Margolis H.S., Bradley D.W.;
RT "The sequence of hepatitis E virus isolated directly from a single
RT source during an outbreak in China.";
RL Virus Res. 28:233-247(1993).
DR EMBL: M94177; AAA96139.1;
DR InterPro: IPR000606; Viral_helicasel.
DR InterPro: IPR002588; V_methyltransf.
DR pfam: PF01443; Viral_helicasel; 1.
DR pfam: PF01660; Methyltransf; 1.
DR SMART: SM01661; DUF27; 1.
SO SEQUENCE 1693 AA; 165271 MW; E33D8F128E0B9C6F CRC64;

```

Query Match 81.0%; Score 7304; DB 12; Length 1693;
 Best Local Similarity 80.8%; Pred. No. 0;
 Matches 1380; Conservative 103; Mismatches 192; Indels 32; Gaps 7;

QY 1 PGTATAEQALAAANSALANAVVBPFLSRVOTELLINIMOPROLVFREVIMNHPIOR 60
 DB 10 PGTATAEQALAAANSALANAVVBPFLSHQOIEILLINIMOPROLVFREVIMNHPIOR 69
 QY 61 VINHELOYCRARAGCLEVGAHPRSRINDNPVNLHRCFLPVRGDOVKTSATRGPAAN 120
 DB 70 VINHELYCRRASGRCLGGAHPRSRINDNPVNLHRCFLPVRGDOVKTSATRGPAAN 129
 QY 121 CRRSALRGPLPAPRTYCEDFGSCAFPAETGVALYSLHDMADVAEAMRHGCTRILYAA 180
 DB 130 CRRSALRGPLPAPRTYCEDFGSCAFPAETGVALYSLHDMADVAEAMRHGCTRILYAA 189
 QY 181 LHLPEVLLPPTGTYHTSTYLLIHDGDAVVTYEGDTSAGYNHDVSLIRAMIRTKIYGDH 240
 DB 190 LHLPEVLLPPTGTYHTSTYLLIHDGDAVVTYEGDTSAGYNHDVSLIRAMIRTKIYGDH 249
 QY 241 PLYIERVRAIGCHFVLLTLTAPEPSPMPYVYPRSTEVYRSTIFGEGSPSLPSSACSTK 300
 DB 250 PLYIERVRAIGCHFVLLTLTAPEPSPMPYVYPRSTEVYRSTIFGEGSPSLPSSACSTK 309
 QY 301 STEHAVPVHIMWRLMLFGATLDDOARCCSRLMTYRGISYKVVGVALYANEGMNASDAL 360
 DB 310 STEHAVPVHIMWRLMLFGATLDDOARCCSRLMTYRGISYKVVGVALYANEGMNASDAL 369
 QY 361 TAXITAVYLTIHQRYLRTQAIKGMRLGVEHAOKFTITLYSLWLFKSGRDYIPGRHQL 420
 DB 370 TAXITAVYLTIHQRYLRTQAIKGMRLGVEHAOKFTITLYSLWLFKSGRDYIPGRHQL 429
 QY 421 FYACRRMTLSAGHLDPRVLVDESVPCRCRTEFLKVAAGFCFPMRLGECCTELEPAE 480
 DB 430 FYACRRMTLSAGHLDPRVLVDESVPCRCRTEFLKVAAGFCFPMRLGECCTELEPAE 489
 QY 481 GLVGDHNDHEAVGSDVPADEPAHLDVSGTYAVHGHQLALYALVPAODIARASRLT 540
 DB 490 GLVGDHNDHEAVGSDVPADEPAHLDVSGTYAVHGHQLALYALVPAODIARASRLT 549
 QY 541 ATVELVASPDLRLECRVILGNKRTFTTVVDAHLEANGPEOYVLSFDAOSMGASHSLT 600
 DB 550 ATVELVASPDLRLECRVILGNKRTFTTVVDAHLEANGPEOYVLSFDAOSMGASHSLT 609
 QY 601 YELTPAGLOYRISNGIDCATATPPPGGAPSAEGEVAFCALYRYNRFTQHSJLJGLM 660
 DB 610 YELTPAGLOYRISNGIDCATATPPPGGAPSAEGEVAFCALYRYNRFTQHSJLJGLM 669
 QY 661 LHPGILGIFPPSPGHIMESANPFCGEGTLTPTWS-TJSGFSSDFSP-----PEAAPA 714
 DB 670 LHPGILGIFPPSPGHIMESANPFCGEGTLTPTWS-TJSGFSSDFSP-----PEAAPA 729
 QY 715 MAATPGLPHSTPRVSDIWLPPSEEPQVDAAPV-PPAPPAGLR-GRVYLTPPPPPVH 772
 DB 730 MAATPGLPHSTPRVSDIWLPPSEEPQVDAAPV-PPAPPAGLR-GRVYLTPPPPPVH 766
 QY 773 KPSTLP-PPSRNRRLTYTPGCAKYAAGSLFESDCDLVYANSNPHRRGGGICLAFYORFP 831
 DB 767 KPSTLP-PPSRNRRLTYTPGCAKYAAGSLFESDCDLVYANSNPHRRGGGICLAFYORFP 826
 QY 832 EAFYPTFEIRREGLAATYTLPRPIIHAVADYREONPKRLLEAAYRETCSSRGGAAPL 891
 DB 827 EAFYPTFEIRREGLAATYTLPRPIIHAVADYREONPKRLLEAAYRETCSSRGGAAPL 886
 QY 892 GSGIYOVPVSLSPDAMERNHRPGDELILTEPANMFEANKPAOAVLTITEDTARTANIAL 951
 DB 887 GSGIYOVPVSLSPDAMERNHRPGDELILTEPANMFEANKPAOAVLTITEDTARTANIAL 946
 QY 952 EIDATATVGRACAGCTISPGIVHQTAFAGVSGSKSRISQOGVDVYVVPTRRLRNSMR 1011
 DB 947 EIDATATVGRACAGCTISPGIVHQTAFAGVSGSKSRISQOGVDVYVVPTRRLRNSMR 1006
 QY 1012 RGFAAFTPHTAARVTIGRRVVIDEAPSLPHLLLHMGRASSVHLLDLPNQIPALIDPEHA 1071

DB 1007 RGFAAFTPHTAARVTIGRRVVIDEAPSLPHLLLHMGRASSVHLLDLPNQIPALIDPEHA 1066
 QY 1072 GLVPAIRPELAPTSWMAVYTHRCPADYCELRGAYPKIOTTSRVLSRFMEHPIAGOKLV 1131
 DB 1067 GLVPAIRPELAPTSWMAVYTHRCPADYCELRGAYPKIOTTSRVLSRFMEHPIAGOKLV 1126
 QY 1132 TOAKKANPGATIVHEAGSTFTETTITATADRGILQSSRAHAYALTHTEKCVTLDA 1191
 DB 1127 TOAKKANPGATIVHEAGSTFTETTITATADRGILQSSRAHAYALTHTEKCVTLDA 1186
 QY 1192 PGLIREVIGISDVIVNNEFLAGGEVGAHPRSVIPRGNDQMLGTLOAFPSCQISAHQOLA 1251
 DB 1187 PGLIREVIGISDVIVNNEFLAGGEVGAHPRSVIPRGNDQMLGTLOAFPSCQISAHQOLA 1246
 QY 1252 EELGHRPAPVAAVLPCCPELEGGILYLPQELTTCDSVTFELTDIVHCRMAAPSQRKAVL 1311
 DB 1247 EELGHRPAPVAAVLPCCPELEGGILYLPQELTTCDSVTFELTDIVHCRMAAPSQRKAVL 1306
 QY 1312 STLVGRTGRTKLYEAAHSDVRESLAFPTTIGPVQATTCCELYEIVAMKEGODGSAYL 1371
 DB 1307 STLVGRTGRTKLYEAAHSDVRESLAFPTTIGPVQATTCCELYEIVAMKEGODGSAYL 1366
 QY 1372 ELDLCNRDVSRIITFFOKXCKFTTGERTIAGKYGOGISAMSKTEFCALFGMPRAIEREIL 1431
 DB 1367 ELDLCNRDVSRIITFFOKXCKFTTGERTIAGKYGOGISAMSKTEFCALFGMPRAIEREIL 1426
 QY 1432 ALLPPIIFYDAYEESVFAAASGAGSCWVFENDESEFSDTONNFSIGLBCVYMECGMP 1491
 DB 1427 ALLPPIIFYDAYEESVFAAASGAGSCWVFENDESEFSDTONNFSIGLBCVYMECGMP 1486
 QY 1492 QWILRLVHLVRSAMIIQAPRESLKGFWKHSRSGEGTLLMNTVMNMAIIACIYERDPRVA 1551
 DB 1487 QWILRLVHLVRSAMIIQAPRESLKGFWKHSRSGEGTLLMNTVMNMAIIACIYERDPRVA 1546
 QY 1552 AFKGDSDVLCSDVROSANAAIILAGGCLKIKYDRIIGYAGVVAAPGLTLPDVVRA 1611
 DB 1547 AFKGDSDVLCSDVROSANAAIILAGGCLKIKYDRIIGYAGVVAAPGLTLPDVVRA 1606
 QY 1612 GRLEKRWGPGPERAEQRLAVCDFTLGLTINVAOVCVVVSRYGVSPGLVHNLIGMLQT 1671
 DB 1607 GRLEKRWGPGPERAEQRLAVCDFTLGLTINVAOVCVVVSRYGVSPGLVHNLIGMLQT 1666
 QY 1672 IADGKAFTETIKPVLDLTNSITQORV 1698
 DB 1667 IADGKAFTETIKPVLDLTNSITQORV 1693

RESULT 14
 069418
 AC 069418; PRELIMINARY; PRT; 1693 AA.
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE ORFL. ORF2 & ORF3.
 OS Hepatitis E virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage.
 OX NCBI_TaxID=12461;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=INDIVIDUAL PATIENT INFECTED WITH HEV;
 RA Von Brunn A., Seebach J., Thyagarajan S.P., Mohanavalli B., Menon T.,
 RA Froesner G.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: X99441; CAA67802.1; -
 DR InterPro: IPR000588; Viral helicase.
 DR InterPro: IPR002589; DUF27.
 DR Pfam: PF01443; Viral_helicase1; 1.
 DR Pfam: PF01660; Vmethyltransf; 1.
 DR Pfam: PF01661; DUF27; 1.
 DR SMART; SM00506; Alpp; 1.

SQ SEQUENCE 1693 AA: 185949 MM: 489B4ADBA5E7E529 CRC64:
 Query Match 81.0%; Score 7301; DB 12; Length 1693;
 Best Local Similarity 79.9%; Pred. No. 0;
 Matches 1367; Conservative 116; Mismatches 188; Indels 40; Gaps 6;

QY 1 PGITTAIEQAALAAANSALANAVVRPLSRVOTELLINMQPRLQVREPLVNMHPICR 60
 DB PGIITVIEQAALAAANSVLANAVVRPLSHOOIEILINMQPRLQVREPLVNMHPICR 69
 QY 61 VHNLEIYOCARAGRCLEAGAHRSINPNVNLHRCFLRPVGRDQVRYMSATRRPAN 120
 DB VHNLEIYOCARAGRCLEAGAHRSINPNVNLHRCFLRPVGRDQVRYMSATRRPAN 129
 QY 121 CRRSALRGPLPADRTYCFDGFSCAFAETGVALYSLHDLMPADVAEAMARHGXLYAA 180
 DB CRRSALRGPLPADRTYCFDGFSCNPAETGVALYSLHDMSPDVAEAMARHGXLYAA 189
 QY 181 LHLPEVLLPFGTYHTTSTYLLHDDGRAVYVEGDTSAGYNHDSILRAMIRTKTIGDH 240
 DB LHLPEVLLPFGTYHTTSTYLLHDDGRAVYVEGDTSAGYNHDSILRAMIRTKTIGDH 249
 QY 241 PLVIEEVRATIGCFEVLTLAAPEPSMPYVPRSTEVYVRSIFGSGSPSLFSPASCSTK 300
 DB PLVIEEVRATIGCFEVLTLAAPEPSMPYVPRSTEVYVRSIFGSGSPSLFSPASCSTK 309
 QY 301 STEHAAPVHIWRLMFLGATLDDQAFCCSRMLTYLRGISYKTVGALVANEGNASEDL 360
 DB STEHAAPVHIWRLMFLGATLDDQAFCCSRMLTYLRGISYKTVGALVANEGNASEDL 369
 QY 361 TAITTAAYLTICHORYLFTQALSKGRRLVEHAQKFTRLXSWLFEKSGRDYIPGRLO 420
 DB TAITTAAYLTICHORYLFTQALSKGRRLVEHAQKFTRLXSWLFEKSGRDYIPGRLO 429
 QY 421 FTAQCRRLISAGHLDPRVLVPRDESPCRCTFLKVKVAKFCFPMWLOECFCELPAB 480
 DB FTAQCRRLISAGHLDPRVLVPRDESPCRCTFLKVKVAKFCFPMWLOECFCELPAB 489
 QY 481 GLVGDGHNEAVEGSEVDPAPBAHLDVSGTYAVHGHOLEALRYALNVODIJAARASRL 540
 DB GLVGDGHNEAVEGSEVDPAPBAHLDVSGTYAVHGHOLEALRYALNVODIJAARASRL 549
 QY 541 AVEELVASPDRLECRCTVLGNKTFRTTVVDGAHLEANGPEBOYVLSFDSASQSMAGSHSLT 600
 DB AVEELVASPDRLECRCTVLGNKTFRTTVVDGAHLEANGPEBOYVLSFDSASQSMAGSHSLT 609
 QY 601 YELTPAGLOVRISNGLDCTATPPPGAPASAPAGEVAACSAIYRYNRTORHSLTGILM 660
 DB YELTPAGLOVRISNGLDCTATPPPGAPASAPAGEVAACSAIYRYNRTORHSLTGILM 669
 QY 661 LHPGGLIGFPPFPGHIMESANPFCGEGTLTRTWS--TSGFSSDPS-----PEAARA 714
 DB LHPGGLIGFPPFPGHIMESANPFCGEGTLTRTWS--TSGFSSDPS-----PEAARA 729
 QY 715 MATPGCLPHSTPVSIDWVLPPESEFQYDAVY--PPAPDPAGLPGVVLTPPPPPVHK 773
 DB MATPGCLPHSTPVSIDWVLPPESEFQYDAVY--PPAPDPAGLPGVVLTPPPPPVHK 782
 QY 774 PSITPP-----SRNRLLYTPDCAKKVAGSLFESDCMLVYNASPGHRPGGGLCHAFY 827
 DB PSITPP-----SRNRLLYTPDCAKKVAGSLFESDCMLVYNASPGHRPGGGLCHAFY 832
 QY 828 OREPEAFYTERIMRGLAAYTLTPRPIIHAAPDYRVONKRLLEAAREFCRSRGTA 887
 DB OREPEAFYTERIMRGLAAYTLTPRPIIHAAPDYRVONKRLLEAAREFCRSRGTA 892
 QY 888 YELLSGTYQVPSLSFDMERNHRCDELVLTLEPAAHFEANKPAQVLTITTEDTART 947
 DB YELLSGTYQVPSLSFDMERNHRCDELVLTLEPAAHFEANKPAQVLTITTEDTART 952
 QY 948 NLALEIDAATENVGRACAGCTISPGIVHVOFTAGVPGSGKSRISIQGGDVVVVPTRELN 1007
 DB NLALEIDAATENVGRACAGCTISPGIVHVOFTAGVPGSGKSRISIQGGDVVVVPTRELN 1012

DB 943 NLALEIDAATENVGRACAGCTISPGIVHVOFTAGVPGSGKSRISIQGGDVVVVPTRELN 1002
 QY 1008 SWRRGFAAFTPHTAARVTIGRRVIDEAPSLPHLLHMQRASSVHLLGDPNQIPAD 1067
 DB 1003 AMRRRGFAFTPHTAARVTIGRRVIDEAPSLPHLLHMQRASSVHLLGDPNQIPAD 1062
 QY 1068 FEHAGLVPAIRDELAPTSMWXYTHRCPADVCELIRGAYKIOGTSRYVLSLFWNEPATIQ 1127
 DB 1063 FEHAGLVPAIRDELAPTSMWXYTHRCPADVCELIRGAYKIOGTSRYVLSLFWNEPATIQ 1122
 QY 1128 KLVYTOAAKAMPGATVTEAGATFETTTIATADARGLIOSRAHVALTRHREKV 1187
 DB 1123 KLVYTOAAKAMPGATVTEAGATFETTTIATADARGLIOSRAHVALTRHREKV 1182
 QY 1188 ILDAPGLREVGISDYIVNPFLLAGGVXGHHPSVLPNGNPONLCTLOAFPSCQISAY 1247
 DB 1183 ILDAPGLREVGISDYIVNPFLLAGGVXGHHPSVLPNGNPONLCTLOAFPSCQISAY 1242
 QY 1248 HOLAEELHRRAPVAALVPCPELEOGLLYMQDELTVSDVLFELTDIVHCRMAAPSOR 1307
 DB 1243 HOLAEELHRRAPVAALVPCPELEOGLLYMQDELTVSDVLFELTDIVHCRMAAPSOR 1302
 QY 1308 KAVLSTLVGRGRRTKLYEAHSDVRESLARIPITGPVQATTCLEYELVEAMVKKGOG 1367
 DB 1303 KAVLSTLVGRGRRTKLYEAHSDVRESLARIPITGPVQATTCLEYELVEAMVKKGOG 1362
 QY 1368 SAVLELIDLCNRPVSRITTFPOKXCNKFTTGETTAHGVGOGISAMSKTEFCALGPMFRATE 1427
 DB 1363 SAVLELIDLCNRPVSRITTFPOKXCNKFTTGETTAHGVGOGISAMSKTEFCALGPMFRATE 1422
 QY 1428 KEILALPPIIFGDAYEESVFAAASGASGCMVEENDESEFDSQNNESLGLCEVMBE 1487
 DB 1423 KEILALPPIIFGDAYEESVFAAASGASGCMVEENDESEFDSQNNESLGLCEVMBE 1482
 QY 1488 CGMPQMLIRLYHLVSAWTLQAPKESIKGEMKKGSEPTLLMNTVMNNAITAHCTEPD 1547
 DB 1483 CGMPQMLIRLYHLVSAWTLQAPKESIKGEMKKGSEPTLLMNTVMNNAITAHCTEPD 1542
 QY 1548 FVVAAFKGDSDVVLCSDYQSRNMAALTAGCCLKLVDRPIGLAGVVAAGLGLTDPV 1607
 DB 1543 FVVAAFKGDSDVVLCSDYQSRNMAALTAGCCLKLVDRPIGLAGVVAAGLGLTDPV 1602
 QY 1608 VRFAGRLSEKMGPPPERAEQDLRLAVCDLRLTVAOVCYDVVSRYGVSPGLVHNLIG 1667
 DB 1603 VRFAGRLSEKMGPPPERAEQDLRLAVCDLRLTVAOVCYDVVSRYGVSPGLVHNLIG 1662
 QY 1668 MLOTTADGKAHTEITIKPYLDJLNSIQRYE 1698
 DB 1663 MLOTTADGKAHTEITIKPYLDJLNSIQRYE 1693

RESULT 15
 Q81873 PRELIMINARY: PRT: 727 AA.
 AC Q81873;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE POLYPROTEIN (ENCODING NTP-BINDING PROTEIN AND RNA-DEPENDENT RNA
 DE POLYMERASE) (FRAGMENT).
 OS Hepatitis E virus.
 OC Viruses: ssRNA positive-strand viruses, no DNA stage.
 OX NCBI_Taxid:12461.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tam A.W., Smith M.M., Kim J.P., Young L.M., Platak M., Feldman R.A.,
 RA Purdy M.A., Bradley D.W., Reyes G.R., Fry K.E.,
 RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=90193694; PubMed=2107574;
 RA Reyes G.R., Purdy M.A., Kim J.P., Luk K.C., Young L.M., Fry K.E.,
 RA Bradley D.;

RT "Isolation of a cDNA from the virus responsible for enterically
transmitted non-A, non-B hepatitis.";
RL Science 247:1335-1339(1990).
RN [3]

RP SEQUENCE FROM N.A.

RX MEDLINE=92271462; PubMed=1589964;

RA Fry K.E., Tam A.W., Smith M.M., Kim J.P., Luk K.C., Young L.M.,

Platak M., Feldman R.A., Yun K.Y., Purdy M.A., Et Al.;

RT "Hepatitis E virus (HEV): strain variation in the nonstructural gene

RT region encoding consensus motifs for an RNA-dependent RNA polymerase

RT and an ATP/GTP binding site.";

RL Virus Genes 6:173-185(1992).

DR EMBL: M32400; AAA03206.1; "

DR InterPro: IPR000606; Viral_helicase1.

DR Pfam: PF01443; Viral_helicase1.

KM Polypeptide: RNA-directed RNA polymerase.

FT NON_TER

SEQUENCE 727 AA; 79306 MW; 872691f6c2318fa7 CRC64;

Query Match 38.2%; Score 3444; DB 12; Length 727;

Best Local Similarity 88.7%; Pred. No. 2.4e-243;

Matches 645; Conservative 40; Mismatches 42; Indels 0; Gaps 0;

QY 972 IVHYQTAGVPGSGSKRSIOGDVYVVPTRRLNSWRGRGFAFPPTAARTIGRRV 1031
DB 1 VVOYQFTAGVPGSGSKRSITQADVDVYVVPTRRLNMRGRGFAFPPTAARTIGRRV 60
QY 1032 VIDAPSLPRLHLLHMQRASSVHLGDPNQIPAIDEFHAGLVPAIRBELAPTSWMXVTH 1091
DB 61 VIDAPSLPRLHLLHMQRAATVHLGDPNQIPAIDEFHAGLVPAIRDLGPTSMWHVTH 120
QY 1092 RCPADVCELTRGATPKIQTTSRVLRSLFWMNEPAIGOKLYTQAKAANPGALITHEAOGA 1151
DB 121 RWPADVCELTRGATPKIQTTSRVLRSLFWMNEPAIGOKLYTQAKAANPGASVTHEAOGA 180
QY 1152 TETETITATADARGLIQSSRAHAIVALTRTEKCVIIDAPGLREVGISDVIYNNFELA 1211
DB 181 TETETITATADARGLIQSSRAHAIVALTRTEKCVIIDAPGLREVGISDVIYNNFELA 240
QY 1212 GGEVGHRRPSYIPRGNDPQNLGLQAFPPSCQISAYHQLABELGHRPAVPAVLPCEPEL 1271
DB 241 GGEIGHRRPSYIPRGNDPQNLGLQAFPPSCQISAFHQLABELGHRPAVPAVLPCEPEL 300
QY 1272 EQGLLYMPQELTSDVLFVELTDIVHCRMAAPSQOKAVLSTLVGRYGRRTKLYEAHSD 1331
DB 301 EQGLLYMPQELTSDVLFVELTDIVHCRMAAPSQOKAVLSTLVGRYGRRTKLYEAHSD 360
QY 1332 VRESLARFIPITIGPVQATTCTELVELVEAMVEKGODGSVLELDLCNRDVSRTFFQKXCN 1391
DB 361 VROSLARFIPITIGPVQATTCTELVELVEAMVEKGODGSVLELDLCNRDVSRTFFQKXCN 420
QY 1392 KFTTGETIANGKVGQGISAMSKTFCALFGPWFRAIEKEITLALPNIIFYGDAYEESVFAA 1451
DB 421 KFTTGETIANGKVGQGISAMSKTFCALFGPWFRAIEKEITLALPQGVFYGDAYEESVFAA 480
QY 1452 AVSAGSCMVFENDSEPDSTQNNFSLGLECVMEECGMPQWLIRLYHLVRSAMILQAPK 1511
DB 481 AVAARAKSMVFENDSEPDSTQNNFSLGLECVMEECGMPQWLIRLYHLVRSAMILQAPK 540
QY 1512 ESLKGFWKHSRGEPGLTLMNTVMNMAITIAHCEFRDPRVAAFKGGDSVILCSDYRQSRNA 1571
DB 541 ESLKGFWKHSRGEPGLTLMNTVMNMAITIAHCEFRDPRVAAFKGGDSVILCSDYRQSRNA 600
QY 1572 AALTAGGLKIKVDYRPIGLAGVVAAGLGLTLPDVVRFAGRLSEKNWGPPEAREQRL 1631
DB 601 AVLAGGLKIKVDYRPIGLAGVVAAGLGLTLPDVVRFAGRLSEKNWGPPEAREQRL 660
QY 1632 AVCDFLGLNVAOVYDVVSRVYGVSPGLVHNILGMLQITADGKAHFTETIKPVLDTN 1691
DB 661 AVSDFLGLNVAOVYDVVSRVYGVSPGLVHNILGMLQITADGKAHFTETIKPVLDTN 720
QY 1692 SIIGRVE 1698

DB 721 SIIGRVE 727

Search completed: March 7, 2002, 14:08:05
Job time: 303 sec
